Computer Aided Characterization of Microcalcification Clusters in Mammograms

Computer Aided Characterization of Microcalcification Clusters in Mammograms

Een wetenschappelijke proeve op het
gebied van de Medische Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
woensdag 4 oktober 2000
des namiddags om 1.30 uur precies

door

Wouter Jan Hendrik Veldkamp

geboren 25 februari 1970 te Voorschoten

Promotor:	Prof. dr. C.C.A.M. Gielen
Co-promotores:	Dr. ir. N. Karssemeijer
	Dr. J.H.C.L. Hendriks
Manuscriptcommissie:	Prof. dr. A.L.M. Verbeek
	Prof. dr. R. Holland
	Dr. ir. J.M. Thijssen

ISBN 90-9013914-1

This research was made possible by Grant KUN 95-897 from
the Dutch Cancer Society

Grants for publication of this thesis by
- the Department of Radiology (UMC St. Radboud, Nijmegen)
- UMC St. Radboud, Nijmegen
- R2 Technology, Inc
- Siemens Nederland N.V.
- Kodak Nederland
are gratefully acknowledged.

Contents

1 Introduction **9**

1.1 Breast diseases 10

1.1.1 Breast anatomy 10

1.1.2 Malignant diseases 11

1.1.3 Benign diseases 12

1.2 Mammography 12

1.2.1 Mammographic abnormalities 13

1.2.2 Conventional and digital mammography 14

1.3 Screening 15

1.4 Computer aided diagnosis 16

1.4.1 Microcalcification clusters 17

1.4.2 Masses, architectural distortions and asymmetric densities 19

1.5 Outline of the thesis 20

Bibliography 22

2 Accurate Segmentation and Contrast Measurement of Microcalcifications in Mammograms: A Phantom Study **29**

2.1 Introduction 29

2.2 Materials and methods 31

2.2.1 Materials 31

2.2.2 Segmentation methods 32

2.2.3 Segmentation and background trend correction 35

2.2.4 Contrast measurement 37

2.2.5 Evaluation methods 37

2.3 Results and discussion 38

2.4 Conclusion 43

2.5 Acknowledgments 44

Bibliography 45

3 Normalization of Local Contrast in Mammograms **47**

3.1 Introduction 47

3.2 Recognition of microcalcifications . . . 52
3.2.1 Feature extraction . . . 52
3.2.2 Statistical model . . . 52
3.3 Optimizing local contrast normalization . . . 53
3.3.1 Determination of bins . . . 53
3.3.2 Asymmetry in histograms of local contrast . . . 55
3.4 Performance evaluation . . . 58
3.5 Results . . . 58
3.6 Discussion and conclusions . . . 61
3.7 Acknowledgments . . . 62
Bibliography . . . 63

4 An Improved Method for Detection of Microcalcification Clusters in Digital Mammograms 65
4.1 Introduction . . . 66
4.2 Detection method . . . 69
4.2.1 Normalization of local contrast . . . 69
4.2.2 Statistical model . . . 70
4.3 Representation of data . . . 71
4.4 Segmentation . . . 71
4.5 Features . . . 72
4.5.1 Individual microcalcification features . . . 73
4.5.2 Cluster features . . . 74
4.6 Feature selection . . . 74
4.7 Classification . . . 75
4.8 Performance evaluation . . . 76
4.9 Results . . . 76
4.10 Discussion and conclusions . . . 80
Bibliography . . . 82

5 Automated Classification of Clustered Microcalcifications into Malignant and Benign Types 85
5.1 Introduction . . . 86
5.2 Materials and Methods . . . 89
5.2.1 Image dataset . . . 89
5.2.2 Image segmentation . . . 89
5.2.3 Microcalcification detection and segmentation . . . 90
5.2.4 Cluster linking . . . 91
5.2.5 Features . . . 92
5.2.6 Feature selection . . . 95

5.2.7 Classification 95
5.2.8 Observer study 96
5.3 Results 97
5.4 Discussion and conclusions 99
Bibliography 102

Summary and conclusions **104**

Samenvatting en conclusies **109**

List of publications **113**

Dankwoord **115**

Curriculum Vitae **117**

Chapter 1

Introduction

In many western countries breast cancer is the most common cancer among women. In the Netherlands approximately 10000 women are diagnosed with the disease annually and about 3500 women die from this type of cancer each year [1]. Over the past decades it has become apparent that breast cancer incidence rates are increasing steadily. Changes in risk factors seem to contribute to the rising incidence [1]. Nevertheless, the mortality rates for breast cancer have remained relatively constant due to more effective treatment and earlier diagnosis [2].

Many risk factors for the disease have been identified, such as family history of breast cancer, late age of first birth, early onset of first menstruation and late age at menopause. As far as current knowledge goes, these risk factors can not explain the major part of the incidence [3]. Moreover, these factors are not a basis for prevention. Therefore, at this point in time, the most important strategy to reduce breast cancer mortality seems early detection through organized breast cancer screening programs. Early diagnosis enables effective treatment and thus increases the survival chance.

Since 1975 pilot projects for breast cancer screening have been carried out in Nijmegen and Utrecht. Currently, the Dutch nation-wide breast cancer screening network invites women aged 50-75 years. Women in this age group are requested once per two years for having a breast examination using mammography (i.e. a technique for making images of the breast using X-rays). Mammography is widely regarded as the most effective method for detection of subtle abnormalities in the breast [4].

Screening for breast cancer is a complex task, due to the small fraction of malignant cases: approximately 5 out of 1000 women in the screening population have breast cancer [5]. To help radiologists in their task to detect signs of cancer between large numbers of normal mammograms, a number of research groups are developing methods for computer aided diagnosis (CAD). Radiologists can use findings of CAD systems as a second opinion and thereby improve their diagnostic performance. Techniques in computer vision and automated pattern recognition are used for this purpose [6].

When a cancerous process in the breast is still at a very early stage, clusters of microcal-

cifications may appear as the only sign. However, 80% of all clusters that are encountered are due to benign processes [7]. In the Netherlands, differentiation of (obvious) benign microcalcification clusters from malignant types during screening is regarded as essential. This is because recalling all cases with microcalcification clusters for further work-up (for instance, using magnification views and biopsies) would result in many false positives (cases that are recalled unnecessary). This would cause a lot of unnecessary anxiety and thereby discourage women to participate. Furthermore, expenses of screening would increase considerably.

Characterization of microcalcification clusters (i.e. differentation between malignant and benign types) is known to be a very difficult task. Moreover, high intra- and inter-observer variabilities have been reported. The objective of the studies described in this thesis was to develop a CAD method for characterization of microcalcification clusters. When a method, as presented in this thesis, can assist radiologists and thereby reduce the number of unnecessarily recalled women with microcalcification clusters, effectiveness of screening could be improved.

1.1 Breast diseases

In this section an overview of important malignant and benign conditions that may appear in the female breast are given. First, we give a short description of the breast anatomy and then malignant and benign breast diseases are discussed respectively.

1.1.1 Breast anatomy

Histologically the breast is divided in 15 to 20 lobes or segments [8, 9, 7]. Working backwards from the nipple, each lobe begins with a major duct that transfers the milk to the nipple during lactation. This major duct branches several times, forming minor (or subsegmental) ducts with correspondingly smaller diameters. Finally, the branching ducts end up in Terminal Ductal Lobular Unit's (TDLU's). A TDLU consists of a lobule and its extralobular terminal duct. The lobule is responsible for milk production during the period of lactation and it is build up from 10 to 100 sac-like units called acini. The ductal and lobular system as a whole is surrounded by an uninterrupted basement membrane on the outside.

With age the breast tissue will change. In young women the breast tissue is dense and rich of glandular tissue. On aging, the glandular tissue is gradually replaced by fat. This increased fat content of the breast in older women makes their mammograms relatively easier to diagnose. This is one of the major reasons that screening programs have excluded women below a certain age (50 years of age in the Dutch screening program). This is reinforced by the need to reduce health risks due to radiation as far as possible [8].

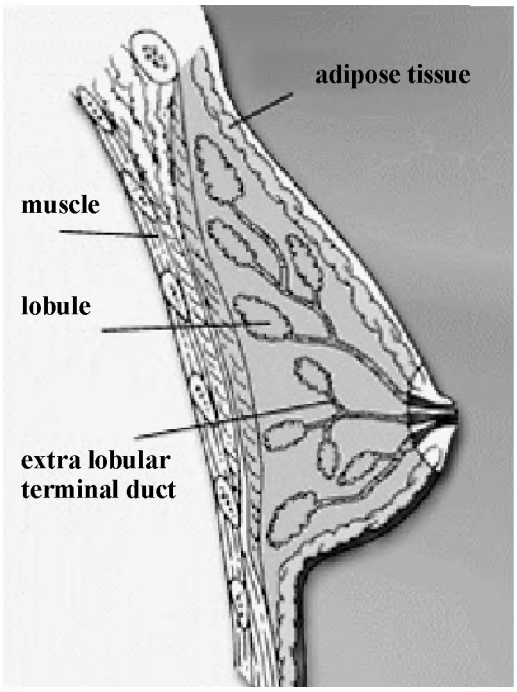


Figure 1.1: Anatomy of the breast (source: www.rmccares.org).

1.1.2 Malignant diseases

Most carcinomas in the breast arise in the TDLU's. The basement membrane plays a key role in determining whether a carcinoma is "in situ" (has not grown through the basement membrane) or "invasive" (has grown through the basement membrane). When in situ carcinomas develop into invasive cancers, they can form metastases to lymph nodes and other organs which will decrease the survival chance. In the breast two forms of in situ malignancy can be considered: the ductal carcinoma in situ (DCIS) and the lobular carcinoma in situ (LCIS). DCIS can involve a variable number of ducts and lobules (in contrast to what the abbreviation DCIS suggests) and often involves central necrosis (i.e. dead cells), but is always contained within the ductal or lobular system. DCIS is considered to be the earliest, detectable stage of breast cancer. The most common progression of DCIS is that it may become an invasive carcinoma. The term DCIS describes a heterogeneous group of lesions [10, 11]. One can distinguish well differentiated, intermediately differentiated, and poorly differentiated DCIS, where the latter type is the most aggressive. In most ductal carcinomas in situ microcalcifications can be found, which may result either from active cellular secretion or from calcified necrosis.

The lobular carcinoma in situ (LCIS) rarely produces microcalcifications and the diagnosis is usually an incidental microscopic finding due to other breast disorders leading to excision of breast tissue. When histologically only LCIS is found no further treatment will be performed. Although LCIS itself forms not a risk of life it is regarded as a risk factor for developing well differentiated DCIS and invasive cancer [9].

1.1.3 Benign diseases

In the following, we will discuss some important benign conditions that show microcalcifications or morphologic growth patterns similar to carcinomas. These conditions are difficult to distinguish from a malignant process. One can consider a number of benign diseases that occur within the lobular and ductal system and that are associated with microcalcification clusters [12, 13, 9]:

- Blunt duct adenosis may be characterized by calcium deposits that form calcifications within the sac-like elements of the lobules (the acini). Typically these are small punctate calcifications which appear uniformly faint or moderate intense on mammograms. The calcifications are separated by fine lines and they are usually collected in round or oval clusters.

- Sclerosing adenosis occurs when proliferation of fibrous tissue causes compression and deformation of lobular structures. The shape of microcalcifications which may occur in the acini, reflects the lobular distortion: they show considerably polymorphism compared to the uniformly punctate calcifications of blunt duct adenosis.

- Microcystic adenosis, also known as cystic hyperplasia, may appear as elongated calcifications resembling the shape of a teacup, seen in oblique mammograms (side view). This effect is caused by milk of calcium that may settle in the cystically dilated acini.

Other important benign conditions occur outside the ductal and lobular system [12, 13, 9]:

- Arterial calcifications of the breast is seen in 3% of mammograms as parallel calcifications of variable lengths.

- Fibroadenomas are benign tumors developing from an over-growth of lobular connective tissue with variable amounts of glandular elements. Fibroadenomas are usually round, oval or lobulated, and can contain coarse and dense microcalcifications.

- Fat necrosis may occur after trauma or inflammation. When fat cells are injured there is a release of fatty acids from the cells. The result is a fibrotic capsule around the oily substance. Calcifications can occur within this capsule and the central oily substance.

1.2 Mammography

Mammography is a diagnostic breast imaging method using X-rays. It is widely used to detect and characterize breast cancer and because of its high performance and low costs it is by far the most suited imaging technique for screening programs.

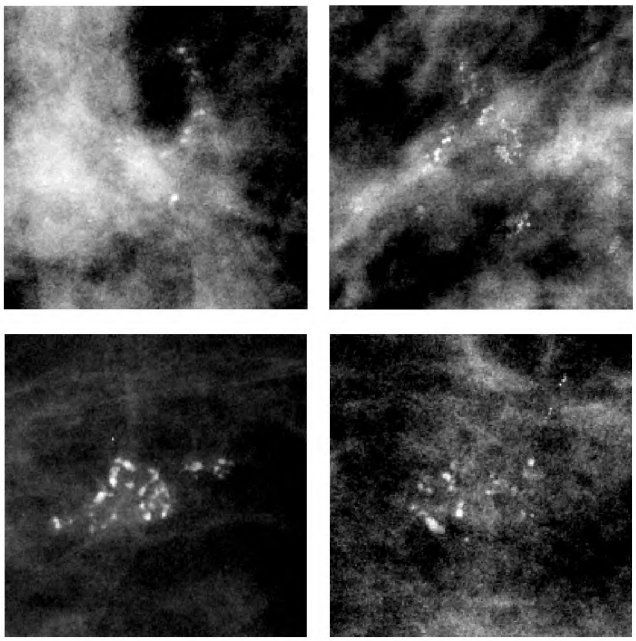


Figure 1.2: Malignant (left) and benign (right) microcalcification clusters.

1.2.1 Mammographic abnormalities

Microcalcification clusters may appear in both in situ and invasive breast cancer but also in benign diseases. Many of the breast cancers that are at an early stage are currently detected by the presence of microcalcifications. Only when appearing as clusters of three or more calcifications, they are clinically suspicious. Microcalcifications that are visible in mammograms vary in diameter roughly from 0.1 *mm* to 0.5 *mm*. Fig. 1.2 shows four clusters. The clusters at the right are benign whereas those presented at the left are malignant. Differentiation between malignant and benign clusters based on mammographic appearance, is not an easy task.

Apart from microcalcification clusters, one can classify the visual signs for which radiologists search during mammographic screening into three basic categories: masses, architectural distortions and asymmetric densities [7]. These abnormalities may indicate invasive breast cancer. Masses that are sharply defined (circumscribed masses) are usually benign. However, if a mass has a faint jagged edge it is likely to be malignant. If a mass is surrounded by a radiating pattern of spicules, it is called a spiculated mass or stellate lesion. Stellate lesions are highly suspicious indicators of breast cancer.

An interruption of the radial ductal pattern is called "architectural distortion". These lesions are often quite subtle and can occur with both benign and malignant processes.

Some masses are detected by radiologists because of asymmetry in the breast pattern between the left and right breast. Asymmetry may be a suspicious sign because in a normal breast the fibro-glandular breast pattern is often symmetric with respect to both breasts.

Figure 1.3: Mammography machine.

1.2.2 Conventional and digital mammography

Fig. 1.3 shows a mammography machine. X-ray photons are emitted from the anode that is located in the X-ray tube on the top of the machine [14]. Part of the radiation goes through the breast and hits a screen. The screen transfers the incident X-ray photons into light photons that blacken the film, which is located just in front of the screen. As a result of interaction between breast tissue and radiation, X-ray photons may undergo a change in direction. This is called scattering which has a blurring effect on the images. By positioning a grid in front of the screen, influence of scatter is reduced. However, using a grid requires a higher radiation dose.

Abnormalities in the breast tissue are often very subtle. Therefore, the mammography machines, film, and developing process are specially designed to create mammograms that are sensitive for these subtle differences. In order to get a good image the breast is compressed, which also enables lowering the dose of radiation. In a standard examination, two images of each breast are taken: one from the top (called a cranio-caudal or *CC* view) and one with the X-ray tube angled approximately 45^0 medially (called a mediolateral oblique or *MLO* view). This ensures that the images display as much breast tissue as possible.

Although conventional (or film-screen) mammography has high sensitivity and specificity, it has some important limitations as well [15]. The film used to capture, store and display the mammographic image is one of the major technical restrictions of film screen mammography. The visibility of breast cancer depends on different attenuation of the X-ray beam by the suspect regions compared with the surrounding tissue. Suspect regions lying in dense areas of the breast may not be noticeable because film contrast decreases in the densest breast areas. This is due to limited dynamic range of conventional films. Furthermore, the image data obtained using a film screen system cannot be manipulated once the image is processed in a film processor. Specifically, over- and under-exposed images have to be recorded again. Contrast levels in the image cannot be altered to improve the relative

Figure 1.4: Mammographic images displayed on two monitors. At the left, a printed film (hard-copy) is shown on a light-box.

visibility of structures in the image without recording additional images of the patient.

Digital mammography has advantages over conventional mammography [15]. Image acquisition and display are decoupled in full field digital mammography. A digital detector is used to capture the X-ray photons. The image is processed by a computer, and can be either printed on film or displayed on monitor (Fig. 1.4). The steps in the breast imaging chain can each be optimized since they are separated. Storage, transmission and retrieval of images can be improved also. Computer aided diagnosis software can be applied to help the radiologist interpret the images. Furthermore, digital detectors have higher dynamic range, and may require a lower X-ray dose.

There are different kinds of digital mammography detectors currently available clinically. For instance, full-field detectors are developed that consist of an array of light-sensitive diodes deposited on a plate of amorphous silicon and covered with a cesium iodide X-ray absorbing phosphor. Other systems use a phosphor screen combined with fiber-optic tapers that transfer the light to charge coupled device (CCD) arrays. Current systems record mammographic images at spatial resolutions of 0.04 and 0.1 *mm* per pixel.

Since four mammographic images (two of each breast) cannot be viewed at full spatial resolution simultaneously on existing monitors on the market today, digital mammograms can be printed to film as an alternative. However, using a film display loses the potential benefits of digital mammography in terms of interactive manipulation of the image. More practical displays will be needed to reach the full potention of digital mammography.

1.3 Screening

In 1989, a national screening program for women between 50 and 70 years of age was started in the Netherlands. In 1998 the upper limit of age was raised to 75 years. In 1990, 8.5% of the breast cancers found in women in the age category 50-69 years was detected by

screening. In 1993 the content of screen-detected carcinomas was increased to 37.2% where 12.9% of these screen-detected carcinomas were in situ carcinomas [5].

It is estimated that the Dutch breast cancer screening program can in the long-term prevent approximately 700 deaths per year from breast cancer. More specifically, a maximum reduction in mortality of 18%, 29% and 23% is expected for women aged 50-59, 60-69 and 70-74 respectively, in the year 2015 [5].

Although screening seems to be effective, recent studies indicate that the radiologic performance can still be improved substantially. It is estimated that in current breast cancer screening programs radiologists do not detect 20 to 25% of the cancers that are visible on retrospective view [16, 17]. Moreover if minimal signs identified on previous screening mammograms are also taken into account, estimates of the number of cancers not reported in screening even increases to 50%, depending on the subjective criteria used by the radiologists performing retrospective reading [18]. A proportion of the missed lesions is likely to be due to inadequate search [19], but it also appears that many of the radiological errors in mammography screening are found to be interpretation errors, i.e. malignant lesions that were consciously judged by a radiologist and reported benign [20].

1.4 Computer aided diagnosis

Computer aided diagnosis (CAD) may be defined as a diagnosis made by a radiologist who takes into account the computer output as a second opinion. CAD aims at improving detection and characterization of abnormalities in mammograms.

Most of the research in computer assisted reading of mammograms focuses at developing methods for detection of abnormalities, like microcalcification clusters, densities and stellate lesions [6]. Successful results are reported in a number of studies [21, 22]. These studies demonstrated a significant increase in radiologist screening efficacy when using CAD. Other research in computer aided diagnosis is focused at helping radiologists in interpreting abnormalities. Impressive results have been reported for computerized characterization of microcalcification clusters and masses [23, 24]. However, it must be noticed that the impact of these methods in a real screening program, where the fraction of abnormal cases is only a few percent (which is totally different in data sets used to evaluate these methods), has yet to be demonstrated.

In this section, we describe some tasks CAD can be used for, and give references to various important approaches and algorithms.

1.4.1 Microcalcification clusters

Segmentation

Shape and contrast features of microcalcifications are often used in characterization schemes for automated classification of microcalcification clusters in malignant and benign types. Also in detection, these features are used to differentiate between true positive and false positive detected microcalcifications. In determining such features, segmentation plays an important role and influences classification or detection performance. By segmentation we mean here determination of the precise outline of a microcalcification, given that it has already been detected. A small number of papers address the problem of optimal segmentation of individual microcalcifications. Most often a thresholding method is used where the threshold level is depending on the image noise or signal level. Levebvre applies a thresholding technique without correcting the image for (low frequency) variations in the background [25]. Other researchers apply thresholding methods on images that are corrected for background trends [26, 27]. Maidment uses thresholding techniques to reconstruct 3D shapes of microcalcifications using stereotactic views [28]. Additional segmentation approaches are described by Parker [29] and Dengler [30].

Detection

A variety of detection methods have been reported in literature. All methods have in common that one or more filters are used to determine local contrast at each pixel inside a region of interest, usually representing the whole breast. Such a local contrast measure is determined by high frequency image noise and local contrast structures occurring in the mammographic image. Detection of microcalcifications is basically searching for high local contrast values in the image.

High frequency image noise limits the detectability of microcalcifications [31, 32]. In order to detect microcalcifications, therefore, reliable measurement of the high frequency image noise is of crucial importance. Most microcalcification detection methods described in literature use some noise dependent adaptive threshold that is locally determined. This is important for dealing with variation of the noise level across the image [33, 21, 34]. Other researchers omit using locally determined thresholds by taking signal dependency of noise into account. In their detection schemes, a pre-processing step of noise equalization is performed which makes the noise level independent of the grey level. Using such an approach in methods for microcalcification detection is found to be beneficial in a number of studies [35, 36, 37, 38]. Additional approaches are based on adaptive noise suppression and other hybrid wavelet adaptive signal enhancement methods [39, 40, 41, 42, 43].

Apart from microcalcifications, other high-contrast structures exist that have high local contrast, such as vessel walls and thin strings of connective tissue. Furthermore, peaks in the high frequency image noise may be hard to distinguish from microcalcifications. Simply

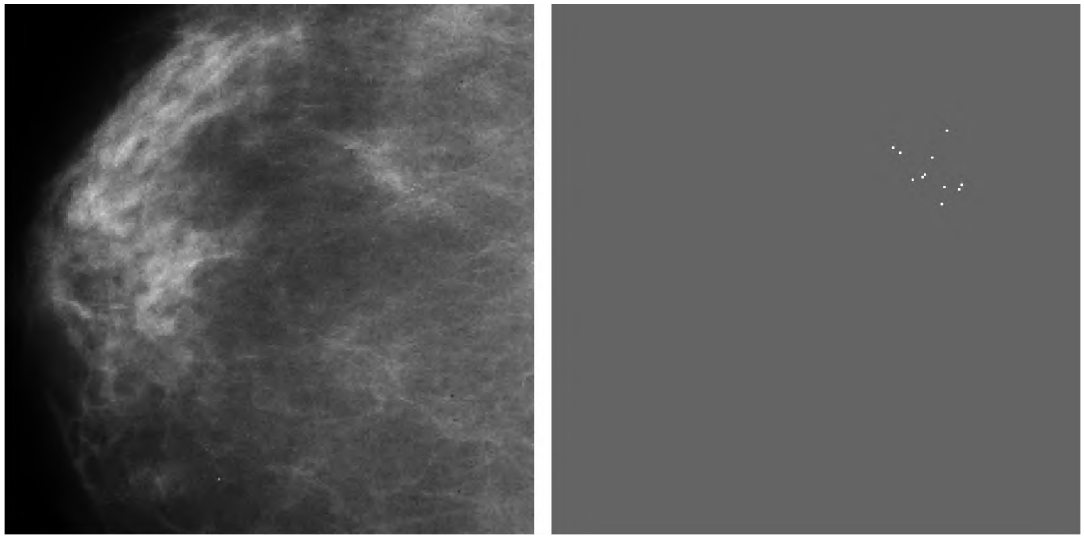


Figure 1.5: Automated detection of microcalcifications. The left image is a cranio caudal view of a breast with a microcalcification cluster. The right-hand image shows the output of an automated detection method developed by Karssemeijer [35]. The white spots represent the microcalcifications the computer found in the mammogram.

selecting pixels with high local contrast yields many false positive findings. Microcalcification detection schemes based on an initial detection step followed by a second step of classification that is intended to remove false positive findings were developed by many researchers [44, 45, 46, 47, 48, 49]. With respect to performance, the current state of art is that the sensitivity of methods for detecting microcalcification clusters can be very high, close to 100% at a false positive rate less than 0.5 per image. Fig. 1.5 shows a *CC* (cranio-caudal) mammogram with microcalcifications. The output image at the right shows a cluster of microcalcifications that was detected by the statistical detection method described in [35].

Characterization

Only few works are related to computer aided characterization (differentiation) of benign and malignant microcalcification clusters. Roughly, features described in literature for characterization schemes are distribution features (based on the distribution of microcalcification properties within a cluster), cluster shape features and texture features. The latter type of features describe texture properties of the local surrounding tissue. Most researchers use either human extracted features ([50, 51]), or manual detection of calcifications along with computer extraction of features as the input for their automatic classification system [27, 52]. Automated characterization applied at microcalcifications that are manually detected by radiologists avoids problems related to detection of false positive or false negative detected microcalcifications when using automated detection. However, such an approach hampers clinical use of automated interpretation of microcalcification clusters. Schmidt addressed this draw-back and he developed a fully automated detection and characterization method

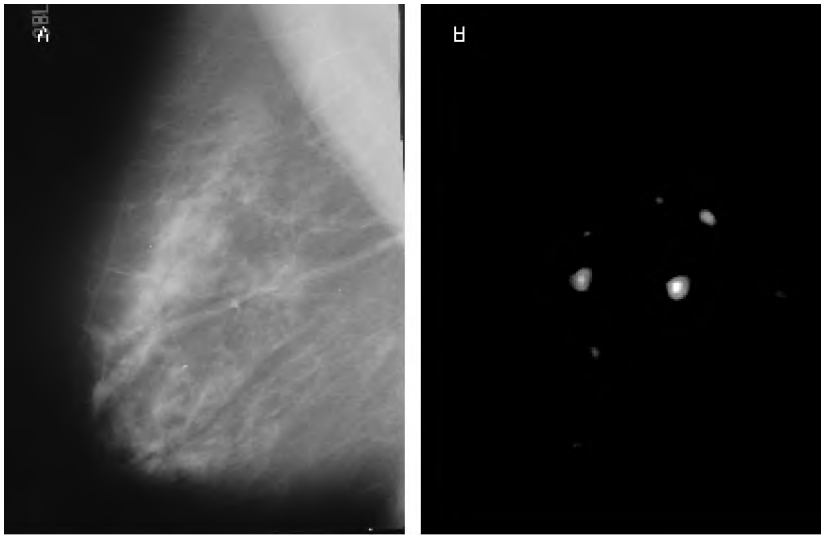


Figure 1.6: A *MLO* image and the corresponding computer output of a stellate lesion that is automatically detected. The intensity level of the spots is a measure for suspiciousness. The brightest spot is at the location of a histologically verified stellate lesion.

for microcalcification clusters [53]. However, he found that his computer system could not infer a reliable diagnosis with respect to cases rated by human experts as being hard to diagnose.

Results reported in literature so far are promising. For instance Jiang compared his method with the mean performance of radiologists in two studies using two different data sets. In both studies his scheme outperformed the radiologists significantly [27, 23]. It must be noticed that this method is not fully automated since microcalcifications are manually identified.

1.4.2 Masses, architectural distortions and asymmetric densities

Segmentation

A large number of segmentation methods have been developed in the field of image analysis and many of them have been used to segment masses in mammograms. One of the most popular segmentation methods, used in many image processing fields, is region growing. Region growing has been applied by a number of groups [54, 55, 56, 57]. Sometimes Markov random fields are involved [58, 54]. Te Brake et al. used a discrete contour model for mass segmentation and they found that their method gave better results than a probabilistic region growing method [59].

Detection

The common approach for detecting lesions suspect for invasive cancer, is generation of target regions by spatial filtering, sometimes involving a measure of asymmetry between

the left and right breast. In a second step the most suspicious regions are selected by a statistical or neural network classifier. Very sensitive algorithms have been developed for masses and stellate lesions, based on analyzing statistical variations of local edge and line orientations [60, 61, 62, 63]. Fig. 1.6 shows an *MLO* (medio-lateral oblique) mammogram and the corresponding output of a CAD system developed by Karssemeijer, where the bright spot is at the location of a histologically verified stellate lesion [64].

Interestingly, good detection results have been reported on cases that were classified retrospectively as missed by screening [17, 65]. It must be noted that the sensitivity for detection of masses, asymmetry and architectural distortions is still lower than for micro-calcification clusters. On the other hand, the number of false positive prompts generated by these methods is steadily decreasing, using the benefits of increased computational power of computers and the availability of huge databases of digital mammograms for training of algorithms.

Characterization

Classification of benign and malignant masses is a well studied subject in mammography. All papers focus on edge analysis of the mass. A vague or spiculated edge indicates malignancy, a sharp well defined contour is likely to belong to a benign abnormality [66, 67, 55]. Additional features that are sometimes used are size, shape, texture and contrast measures.

1.5 Outline of the thesis

The objective of this thesis was to develop a fully automated method for characterization of microcalcifications. By characterization we mean classification in malignant and benign types. Such a method consists of two main steps:

1. Automated detection of microcalcification clusters

2. Automated characterization of (true positive detected) microcalcification clusters.

In both these steps, algorithms for automated segmentation of microcalcifications are often involved.

Before we describe methods for automated detection and characterization we focus in **Chapter 2** on methods for segmentation of microcalcifications. In this chapter a number of microcalcification segmentation methods are investigated in a phantom study. For performing a classification specific characteristics or features of microcalcifications have to be determined. Therefore, finding the precise outline of the calcifications is important. Mammographic recordings of the The CDMAM phantom, that consists of a pattern of dots with known size and object contrast, are used for evaluation of contrast measurement and

segmentation. Dots in the range of 0.2 *mm* to 0.8 *mm* are taken as a model for microcalcifications. Three methods based on thresholding and one iterative method based on a Markov random field are investigated. The performances of the different methods for segmentation are compared. Finally, the influence of the Modulation Transfer Function (MTF) on contrast estimates is determined and the effect of exposure level on segmentation is analyzed.

Previously, a microcalcification detection method was developed by Karssemeijer [31]. An important step in his detection scheme is a noise equalization step that compensates for signal dependency of high frequency image noise.

In **Chapter 3**, the effect on detection performance of new approaches that correct for signal dependency of high frequency noise is investigated. For this purpose, noise characteristics are estimated from the image itself without using additional information obtained from phantom recordings. This makes the approach robust and independent of film type and film development characteristics. Furthermore, it should be possible to apply the method on direct digital mammograms as well.

In **Chapter 4**, an effort is made to improve the detection performance by using a second step of classification. This second step aims at filtering out false positive detected clusters. For classification the *k*-nearest neighbor method is used in a leave-one-patient-out procedure. A feature selection procedure is used for selection of useful features. By applying the initial detection at various levels of sensitivity, various sets of false and true positive detected clusters are created. At each of these sets the classification can be performed. It is investigated whether it is beneficial to perform classification at sets obtained at lower (initial) detection sensitivity.

In **Chapter 5**, a characterization method is described for performing the final step of classifying microcalcification clusters into malignant and benign types. The objectives in this study are to design and test a fully automated method for classification of microcalcification clusters into malignant and benign types, and to compare the method's performance with that of radiologists. Microcalcifications are detected automatically in cluster regions that are annotated according to radiologic screening reports. Specific features with respect to the detected candidate microcalcifications within the marked area are calculated and used for classification. Novel aspects of the approach are that the relative location and orientation of clusters inside the breast are taken into account for feature calculation. Furthermore, correspondence of location of clusters in *MLO* and *CC* views, is used in feature calculation and in final classification. A total of sixteen features is used in the study. For classification the *k*-nearest-neighbor method is used in a leave-one-patient-out procedure.

The data set that is used in this study does not contain obvious benign cases. A subset of the data set, containing mammograms from 90 patients, is used for comparing the computer results to human performance. These mammograms are read on a light-box by ten radiologists and the probability of malignancy is assessed for each patient.

Bibliography

[1] Signaleringscommissie kanker van de Nederlandse kankerbestrijding/KWF. Signaleringsrapport kanker 1999. Sophialaan 8, 1075 BR Amsterdam, The Netherlands, 1999.

[2] M J M Broeders and A L M Verbeek. Breast cancer epidemiology and risk factors. *Q-J-Nucl-Med*, 41(3):179–88, 1997.

[3] C H van Gils. *Mammographic density and breast cancer risk*. PhD thesis, Katholieke Universiteit Nijmegen, the Netherlands, 1998. ISBN 90-9011983-3.

[4] D Brzakovic and M Neskovic. Mammogram screening using multiresolution-based image segmentation. In K W Bowyer and S M Astley, editors, *State of the art in digital mammographic image analysis*, volume 9 of *Series in machine perception and artificial intelligence*, pages 103–127. World Scientific, 1994.

[5] Landelijk evaluatie team voor bevolkingsonderzoek naar borstkanker. Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in nederland. Instituut voor maatschappelijke gezondheidszorg, Erasmus Universiteit Rotterdam, Postbus 1738, 3000 DR Rotterdam 1999, May 1999.

[6] N Karssemeijer and J H C L Hendriks. Computer-assisted reading of mammograms. *European radiology*, 7(5):743–748, 1997.

[7] L Tabar. Teaching cours of mammography. diagnosis and in-depth differential diagnosis of breast cancer. Mammograhy Education, Inc., Dept. of Radiology, University of Uppsala School of Medicine, Sweden.

[8] R Highnam and M Brady. *Mammographic image analysis*. Kluwer Academic Publishers, P.O. Box 17, 3300 AA Dordrecht, The Netherlands, 1 edition, 1999.

[9] D Dronkers, J H C L Hendriks, R Holland, and G Rosenbusch. *Radiologische mammadiagnostik*. Georg Thieme Verlag, Rüdigerstrasse 14, D-70469 Stuttgart, 1999.

[10] J R Harris, S H Hellman, I C Henderson, and D W Kinne. *Breast diseases*. J.B. Lippincott Company, East Washington Square, Philadelphia, Pennsylvania 19105, 1991.

[11] R Holland, J L Peterse, R R Millis, V Eusebi, D Faverly, M J van de Vijver, and B Zafrani. Ductal carcinoma in situ: a proposal for a new classification. *Seminars in diagnostic pathology*, 11(3):167–180, 1994.

[12] C Daniels, G H Schaller, N L McNeil, P Salvador, T Matcalfe, J S Caines, and N A Wadden. Fundamentals of breast imaging, a multimedia teaching file for residents, technologists, radiologists and physicists. Radiology dept. Dalhousie University, Halifax, Nova Scotia; Radiology dept. University of Toronto, 1997. ISBN: 0-7703-2806-7.

[13] M Lanyi. *Breast calcifications.* Springer-Verlag, Berlin Heidelberg, 1988.

[14] T S Curry, J E Dowdey, and R C Murry jr. *Christensen's physics of diagnostic radiology.* Lea & Febiger, 200 Chester Field Parkway, Malvern, Pennsylvania 19355-9725, USA, 4 edition, 1990.

[15] ED Pisano. Initial clinical experience with full field digital mammography. In N Karssemeijer, MAO Thijssen, JHCL Hendriks, and LJTO van Erning, editors, *Digital Mammography*, pages 391–394. Kluwer, Dordrecht, 1998.

[16] R G Blanks, M G Wallis, and S M Moss. A comparison of cancer detection rates achieved by breast cancer screening programmes by number of readers, for one and two view mammography: results from the uk national health breast screening programme. *J Med Screen*, 5:195–201, 1998.

[17] L J Warren Burhenne, S A Wood, C J D'Orsi, S A Feig, D B Kopans, K F O'Shaugnessy, E A Sickles, L Tabar, C J Vyborny, and R A Castelino. The potential contribution of computer aided detection to the sensitivity of screening mammography. *Radiology*, 2000.

[18] J E Harvey, L L Fajardo, and C A Inis. Previous mammograms in patients with impalpable breast carcinoma: retrospective vs blinded interpretation. *AJR*, 161:1167–1172, 1993.

[19] C J Vyborny. Can computers help radiologists read mammograms? *Radiology*, 191:315–317, 1994.

[20] M Hartswood, R Procter, and LJ Williams. Prompting in practice: How can we ensure radiologists make best use of computer-aided detection systems in screening mammography. In N Karssemeijer, MAO Thijssen, JHCL Hendriks, and LJTO van Erning, editors, *Digital Mammography*, pages 363–370. Kluwer, Dordrecht, 1998.

[21] H P Chan, K Doi, C J Vyborny, R A Schmidt, C E Metz, K L Lam, T Ogura, Y Wu, and H Macmahon. Improvement in radiologist's detection of clustered microcalcifications on mammograms. *Inv Radiol*, 25:1102–1110, 1990.

[22] W P Kegelmeyer, J M Pruneda, P D Bourland, A Hillis, M W Riggs, and M L Nipper. Computer-aided mammographic screening for spiculated lesions. *Radiology*, 191:331–337, 1994.

[23] Y L Jiang, R M Nishikawa, R A Schmidt, C E Metz, M L Giger, and K Doi. Improving breast cancer diagnosis with computer-aided diagnosis. *Acad Radiology*, 6(1):22–33, 1999.

[24] H P Chan, B Sahiner, M A Helvie, N Petrick, M A Roubidoux, T E Wilson, D D Adler, C Paramagul, J S Newman, and S Sanjay-Gopal. Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: An roc study. *Radiology*, 212:817–827, 1999.

[25] F Levebvre, R Gilles, J Masselot, and E Kahn. Computer characterization of clustered microcalcifications in mammograms. In H U Lumke, M L Rhodes, C C Jaffe, and R Felix (Springer-Verlag Berlin 1991), editors, *Computer Assisted Radiology Proceedings of the International Symposium*, 1991.

[26] H P Chan, Lo SCB, L T Niklason, D M Ikeda, and K L Lam. Image compression in digital mammography: Effects on computerized detection of subtle microcalcifications. *Medical physics*, 23(8):1325–1336, 1996.

[27] Y L Jiang, R M Nishikawa, D E Wolverton, C E Metz, M L Giger, R A Schmidt, C J Vyborny, and K Doi. Malignant and benign clustered microcalcifications: Automated feature analysis and classification. *Radiology*, 198:671–678, 1996.

[28] Maidment ADA, E F Conant, S A Feig, C W Piccoli, and M Albert. 3-dimensional analysis of breast calcifications. In K Doi, M L Giger, R M Nishikawa, and R A Schmidt, editors, *Digital Mammography*, pages 245–250. Elsevier, Amsterdam, 1996.

[29] J Parker and D R Dance J Suckling. The hysteresis technique for detection of calcifications in digital mammograms,. In A G Gale, S M Astley, D R Dance, and A Y Cairns, editors, *Digital Mammography*, pages 41–50. Elsevier, Amsterdam, 1994.

[30] J Dengler, S Behrens, and J.F. Desaga. Segmentation of microcalcifications in mammograms. *IEEE Trans. on Medical Imaging*, 12(4):634–642, May 1993.

[31] N Karssemeijer, J T Frieling, and J H Hendriks. Spatial resolution in digital mammography. *Invest radiol*, 28(5):413–9, May 1993.

[32] H Roehrig, E Krupinski, and T Yu. Physical and psychophysical evaluation of digital systems for mammography. *SPIE 2436*, pages 124–134, 1995.

[33] R M Nishikawa, M L Giger, K Doi, C J Vyborny, and R A Schmidt. Computer-aided detection of clustered microcalcifications: an improved method for grouping detected signals. *Med Phys*, 20(6):1661–1666, 1993.

[34] W P Kegelmeyer and M C Allmen. Dense feature maps for detection of calcifications. In A G Gale, S M Astley, D R Dance, and A Y Cairns, editors, *Digital Mammography*, pages 3–12. Elsevier, Amsterdam, 1994.

[35] N Karssemeijer. A stochastic model for automated detection of calcifications in digital mammograms. *Image and Vision Computing*, 10:369–375, 1992.

[36] T Netsch and H-O Peitgen. Scale space signatures for the detection of clustered microcalcifications in digital mammograms. *to appear in IEEE Transactions on Medical Imaging*, 18:774–786, 1999.

[37] A Maitournam, C Graffigne, and A Strauss. Modeling of digital mammograms using bicubic spline functions and additive noise. In N Karssemeijer, MAO Thijssen, JHCL Hendriks, and LJTO van Erning, editors, *Digital Mammography*, pages 177–184. Kluwer, Dordrecht, 1998.

[38] S Brown, R Li, L Brandt, L Wilson, G Kossoff, and M Kossoff. Development of a multi-feature cad system for mammography. In N Karssemeijer, MAO Thijssen, JHCL Hendriks, and LJTO van Erning, editors, *Digital Mammography*, pages 189–196. Kluwer, Dordrecht, 1998.

[39] R N Strickland and H I Hahn. Wavelet transforms for detecting microcalcifications in mammograms. *IEEE Transactions on Medical Imaging*, 15(2):218–228, 1996.

[40] W Qian, M Kallergi, L P Clarke, H D Li, P Venugopal, D Song, and R A Clark. Tree structured wavelet transform segmentation of microcalcifications in digital mammography. *Med phys*, 22(8):1247–54, Aug 1995.

[41] W Qian, L P Clarke, M Kallergi, and R A Clark. Tree-structured nonlinear filters in digital mammography. *IEEE Trans. Med. Imag.*, 13:25–36, 1994.

[42] W Richardson. Nonlinear filtering and multiscale texture discrimination for mammograms. *Proc SPIE 1768, Mathematical Methods Med. Imag.*, pages 293–305, 1992.

[43] A F Laine and S Song. Multiscale wavelet representation for mammographic feature analysis. *Proc SPIE 1768, Mathematical Methods Med. Imag.*, pages 306–316, 1992.

[44] R M Nishikawa, Y Jiang, and M L Giger et al. Performance of automated cad schemes for the detection and classification of clustered microcalcifications. In A G Gale, S M Astley, D R Dance, and A Y Cairn, editors, *Digital Mammography*, pages 13–20. Elsevier, Amsterdam, 1994.

[45] C S Carman and G Eliot. Detecting calcifications and calcification clusters in digitized mammograms. In K Doi, M L Giger, R M Nishikawa, and R A Schmidt, editors, *Digital Mammography*, pages 253–255. Elsevier, Amsterdam, 1996.

[46] T Hara, K Hirako, H Fujita, T Endo, K Horita, M Ikeda, C Kido, and T Ishigaki. Automated detection algorithm for clustered microcalcifications based on density gradient

and triple-ring filter analysis. In K Doi, M L Giger, R M Nishikawa, and R A Schmidt, editors, *Digital Mammography*, pages 257–262. Elsevier, Amsterdam, 1996.

[47] P Sajda, C D Spence, J C Pearson, and R M Nishikawa. Integrating multi-resolution and contextual information for improved microcalcification detection in cad. In K Doi, M L Giger, R M Nishikawa, and R A Schmidt, editors, *Digital Mammography*, pages 291–296. Elsevier, Amsterdam, 1996.

[48] D Meersman, P Scheunders, and D van Dyck. Classification of microcalcifications using texture-based features. In N Karssemeijer, M A O Thijssen, J H C L Hendriks, and L J T O van Erning, editors, *Digital Mammography*, pages 233–236. Kluwer, Dordrecht, 1998.

[49] R F Nagel, R M Nishikawa, J Papaioannou, and K Doi. Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms. *Medical physics*, 25(8):1502–1506, 1998.

[50] J A Baker, P J Korngruth, Y J Lo, M E Williford, and C E Floyd jr. Breast cancer: prediction with artificial neural network based on bi-rads standardized lexicon. *Radiology*, 196:817–820, 1995.

[51] Y Wu, M L Giger, K Doi, C J Vyborny, R A Schmidt, and C E Metz. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology*, 187:81–87, 1993.

[52] H P Chan, B Sahiner, K L Lam, N Petrick, M A Helvie, M M Goodsitt, and D D Adler. Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces. *Medical physics*, 25(10):2007–2019, October 1998.

[53] F Schmidt, E Sorantin, C Szepesvari, E Graif, M Becker, H Mayer, and K Hartwagner. An automatic method for the identification and interpretation of clustered microcalcifications in mammograms. *Phys. med. biol*, 44:1231–1243, 1999.

[54] H D Li, M Kallergi, L P Clarke, V K Jain, and R A Clark. Markov random field for tumor detection in digital mammography. *IEEE Trans on Med Imag*, 14:565–576, 1995.

[55] S Pohlman, K A Powell, N A Obuchowski, W A Chilcote, and S Grundfest-Broniatowski. Quantitative classification of breast tumors in digitized mammograms. *Med Phys*, 23:1337–1345, 1996.

[56] A J Mendez, P G Tahoces, M J Lado, M Souto, and J J Vidal. Computer-aided diagnosis: automatic detection of malignant masses in digitized mammograms. *Medical Physics*, 25(6):957–964, June 1998.

[57] MA Kupinski and ML Giger. Automated seeded lesion segmentation on digital mammograms. *IEEE Trans on Med Imaging*, 17:510–517, 1998.

[58] M L Comer, S Liu, and E J Delp. Statistical segmentation of mammograms. In K Doi, M L Giger, R M Nishikawa, and R A Schmidt, editors, *Digital Mammography*, pages 475–478. Elsevier, Amsterdam, 1996.

[59] G M te Brake, M J Stoutjesdijk, and N Karssemeijer. A discrete dynamic countour model for mass segmentation in digital mammograms. *SPIE Medical Imaging 1999*, 3661:911–919, 1999.

[60] K Woods and K Bowyer. A general view of detection algorithms. In K Doi, M L Giger, R M Nishikawa, and R A Schmidt, editors, *Digital Mammography*, pages 385–390. Elsevier, Amsterdam, 1996.

[61] M Zhang and M L Giger. Automated detection of spiculated lesions and architectural distortions in digitized mammograms. *SPIE 2434*, pages 846–855, 1995.

[62] T C Parr, S M Astley, C J Taylor, and Boggis CRM. Model based classification of linear structures in digital mammograms. In K Doi, M L Giger, R M Nishikawa, and R A Schmidt, editors, *Digital Mammography*, pages 351–356. Elsevier, Amsterdam, 1996.

[63] N Karssemeijer and G M te Brake. Detection of stellate distortions in mammograms. *IEEE Trans Med Imag*, 15:611–619, 10 1996.

[64] N Karssemeijer. Detection of stellate distortions in mammograms using scale space operators. In Y Bizais, C Barrilot, and R Di Paola, editors, *Information Processing in Medical Imaging*, pages 335–346. Kluwer, Dordrecht, 1995.

[65] G M te Brake and N Karssemeijer. Automated detection of breast carcinomas that were not detected in a screening program. *Radiology*, 207:465–471, 1998.

[66] M L Giger, C J Vyborny, and R A Schmidt. Computerized characterization of mammographic masses : analysis of spiculation. *Cancer Letters*, 77:201–211, 1994.

[67] Z Huo, M L Giger, C J Vyborny, U Bick, P Lu, D E Wolverton, and R A Schmidt. Analysis of spiculation in the computerized classification of mammographic masses. *Med Phys*, 22:1569–1579, 10 1995.

Chapter 2

Accurate Segmentation and Contrast Measurement of Microcalcifications in Mammograms: A Phantom Study[1]

ABSTRACT

We are developing a computer aided diagnostic method to assist radiologists in differentiating between malignant and benign clustered microcalcifications in mammograms. In earlier studies we investigated shape and contrast features of microcalcifications for classification. It was found that segmentation strongly influences classification results. For this reason a phantom study has been carried out. The CDMAM phantom, consisting of a pattern of dots with known size and object contrast is used for evaluation of contrast measurement and segmentation. Dots in the range of 0.2 *mm* to 0.8 *mm* are taken as a model for microcalcifications. In this article performances of different methods for segmentation of microcalcifications are compared. An iterative method based on a Markov random field and a signal dependent criterion give satisfying results. The segmentation performances of both methods are comparable. Also the influence of the Modulation Transfer Function (MTF) on contrast estimates is determined and effect of exposure level on segmentation is analyzed.

2.1 Introduction

Shape and contrast features of microcalcifications are often used in schemes for automated classification of microcalcification clusters in malignant and benign types. Also in detection, these features are used to differentiate between true positive and false positive detected microcalcifications. In determining such features, segmentation plays an important role and influences classification or detection performance. By segmentation we mean here

[1]This chapter is based on the publication: W.J.H. Veldkamp, N. Karssemeijer, *Accurate segmentation and contrast measurement of microcalcifications in mammograms: A phantom study*, Medical Physics, vol 25, nr.7, 1998.

determination of the precise outline of a microcalcification given that it has already been detected. A variety of detection and classification methods have been reported in literature [1, 2, 3, 4, 5, 6], but only few papers address the problem of optimal segmentation of individual microcalcifications. H.P. Chan et al. use a local noise based threshold for region growing with background trend correction [7]. F. Levebvre et al. use a local noise based threshold for region growing in a signal enhanced image without background trend correction [8]. Y. Jiang et al. apply a background trend correction and use a signal dependent threshold for region growing [9]. Other approaches are described by Parker [10] and Dengler [11]. In this phantom study we analyzed performances of different segmentation methods.

Determination of microcalcification contrast relies on accurate segmentation. Overestimation (or underestimation) of object size deteriorates contrast measurement significantly. Apart from a proper segmentation method, also influence of MTF has to be taken into account in contrast estimation. Furthermore, it is important to define contrast in a way that influence of breast thickness, exposure level and digitization is eliminated. We are interested in determination of the average linear attenuation coefficient of microcalcifications, because potentially this may be an important feature for classification of microcalcification clusters into benign and malignant types. An effort can be made to determine the attenuation factor of microcalcifications from their contrast and thickness when we neglect the linear attenuation coefficient of the background. This is true in good approximation since the attenuation factor of the latter is much smaller [12]. Estimation of microcalcification thickness will require some assumptions about 3D structure. More accurate estimation is possible when multiple views are available [13].

In previous studie we found that segmentation strongly influences results of classifying microcalcifications into malignant or benign types using shape and contrast features [14]. Therefore, in this article, we investigate methods for automated segmentation and contrast measurement of microcalcifications. However, a ground truth for segmentation of microcalcifications in mammograms is hardly obtainable since annotation is hampered by mammographic noise. For this reason we used a phantom that consists of 205 square cells with dots of varying size and contrast for evaluation of segmentation and contrast measurement (see Fig. 2.1). A computer program is developed that performs an automated read-out of the phantom recordings and is used in the segmentation scheme for automated detection of the dots. Segmentation is performed at the location of the marks that indicate the presence of a dot. A number of segmentation methods is investigated. Three methods use a threshold value that is locally determined from the image data. The fourth approach is an iterative method based on a Markov random field model, with parameters estimated locally. Experiments with a mammographic background added to the phantom images are carried out to investigate use of background trend correction on segmentation. We applied a two-step procedure here. A low-pass filtered background image is obtained and subtracted from the original image in a first step. The final segmentation is performed on the corrected image.

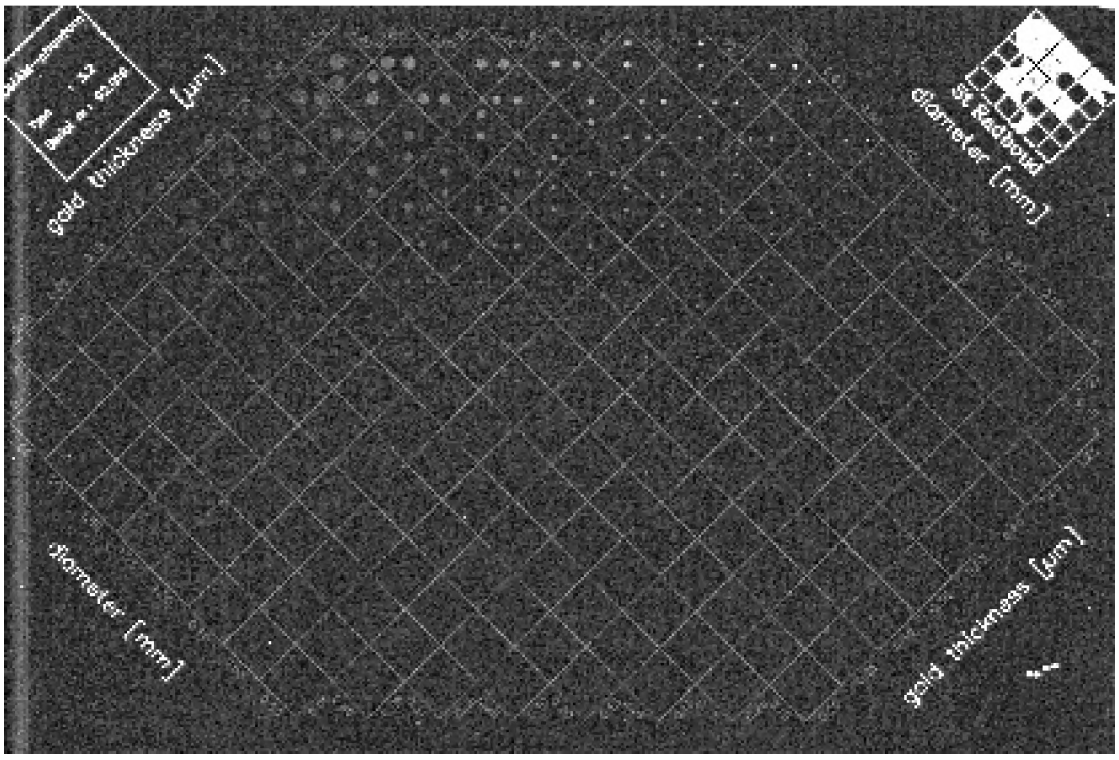


Figure 2.1: An X-ray image of the CDMAM phantom.

Using an accurate segmentation method the influence of the MTF (Modulation Transfer Function) of the imaging system on contrast measurement is determined. Finally, segmentation results from recordings obtained at three different exposure levels are compared.

2.2 Materials and methods

2.2.1 Materials

The phantom we used is the CDMAM phantom that was developed at our institute and especially designed for quality assurance in mammography. This phantom consists of a matrix of square cells with golden dots varying in size and object contrast. In each of the 205 cells of the matrix one dot is at the center and another is positioned in a randomly selected corner (Fig. 2.1). Dot diameter is varying from 3.20 *mm* to 0.10 *mm* and contrast from 0.05 μm to 1.60 μm gold thickness. The golden dots are damped onto an aluminum layer of 0.5 *mm*. The phantom was recorded at 3 exposure levels (optical density OD = 0.55, 1.21 and 1.84) with the mammomat 3 (SIEMENS, focalspot 0.1-0.3; molybdenum anode and filter) where the phantom was placed within 4 *cm* perspex. The conventional images were recorded with a mammographic film/screen system (KODAK MIN-RH/MIN-R). At every exposure level 8 recordings were made. The recordings were digitized at 50 micron per pixel (12 bits pixels) with a Lumisys 85 digitizer.

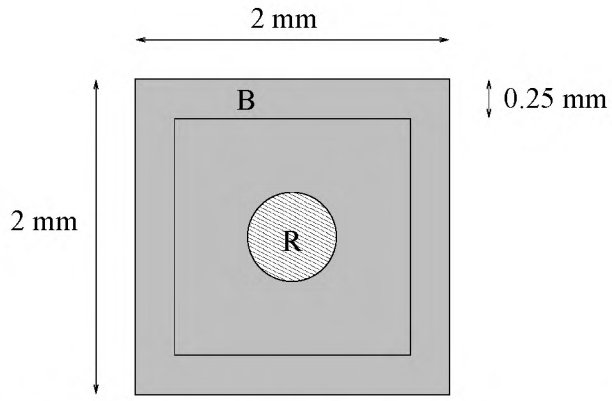


Figure 2.2: Schematic drawing. Each dot with its local background is stored in a 40×40 pixel data field (50 micron per pixel), where its center of mass coincides with the data fields center.

2.2.2 Segmentation methods

In [15] a method is described for computation of contrast detail curves by automated read-out of phantom images and calculation of the probability of detection for each cell. This program was used to determine the location of each dot, exact in the digital image matrix. In our segmentation experiments we used all dots that have a probability of detection $p(d) \geq 90\%$. For segmentation experiments each dot with its local background is stored in 40×40 pixel data fields (50 micron per pixel), where its center of mass coincides with the data fields center (Fig. 2.2). The segmentation program performs a segmentation on each of the data fields. It is important to note that the original annotation, given by the detection program, is not involved in the segmentation procedure. It is assumed, however, that the center of the data field is part of the detected dot. A number of segmentation methods is investigated. A distinction is made between thresholding and an iterative approach.

Thresholding Methods

Thresholding is a well-known tool in image segmentation. Let y_i denote the intensity of a pixel at site i and let ${y_i}^T$ denote the value of a pixel at site i in a binary image. Then thresholding y_i at grey level T will result in ${y_i}^T$ as follows [16]

$$ {y_i}^T = \begin{cases} 0 & \text{if } y_i < T \\ 1 & \text{if } y_i \geq T \end{cases} \tag{2.1} $$

With 0 and 1 a pair of binary gray levels. Global and local thresholding techniques can be considered. A global technique applies a single threshold to the entire image, whereas in a local technique thresholds are determined for each pixel or for small sub-images. Three segmentation methods based on thresholding are considered here.

1. A local noise dependent threshold (LNDT)
2. A fixed contrast threshold (FCT)
3. A local signal intensity dependent threshold (LSDT)

Thresholds are determined for each data field independently and one object is assumed per field. For all methods, the mean background level and background noise are estimated from the outer rim of 5 pixels in width of the 40×40 pixel data field. This region will be indicated by B and it only contains background pixels, since all objects under study are smaller than 20 pixels in width. After thresholding, regions of marked pixels ($y_i{}^T = 1$) within the 40×40 data field are determined. The 4-connected region in the center of the data field is marked as the segmented object (indicated by R).

Local noise dependent threshold

When the background region contains a lot of noise, regions of thresholded background pixels may be linked to the segmented object. To prevent thresholding noisy pixels in the neighborhood of a microcalcification it might be beneficial to increase the segmentation threshold with the noise level in the background.

High frequency noise n in the background region B can be estimated by calculating local contrast c_i of pixels at site i [2]:

$$c_i = y_i - \frac{1}{N}\sum_{j\in\partial_i} y_j \qquad for\ i \in B, \tag{2.2}$$

with ∂_i a neighborhood or window at i of size N. Assuming the probability density of local contrast to be symmetrical around zero, the expected value $E(|c|)$ is proportional to σ_c, the standard deviation of c. In our implementation a measure for local high frequency noise n is defined by

$$n = \frac{1}{M}\sum_{i=0}^{M-1} |c_i|, \tag{2.3}$$

where M is the number of pixels in B. A pixel is marked if its value exceeds a certain noise dependent level T_n.

$$T_n = y_B + a_n n. \tag{2.4}$$

With y_B the mean pixel value of the background and a_n is a constant.

Fixed contrast threshold

The segmented size of an object will depend on the threshold level. This is obvious for

spherical objects, but also for cylindrical objects like the phantom dots this will be the case, since sharp edges in a signal are blurred by the point spread function (PSF) of the imaging system. If this effect is significant it can cause inconsistency in segmenting objects of equal size and contrast when the local noise dependent criterion is applied. To aim at more consistency, a fixed threshold can be used instead of adapting the threshold to the noise level. In the fixed contrast method pixels are thresholded when their contrast to the background exceeds a certain level. Pixel contrast is calculated by taking the difference in pixel intensity and the mean background intensity.

After digitization and using a logarithmic scaling, pixel values y_i are related to optical density OD by

$$y_i = c_0[\log I_0 - OD], \tag{2.5}$$

with I_0 and c_0 constants related to calibration of the film digitizer. For estimation of microcalcification contrast we use a method reported in [17]. In first approximation and ignoring the MTF of the imaging system, an expression for microcalcification pixel contrast can be derived [14]. This contrast measure C_i^{mc} depends on microcalcification thickness d_{mc}, an estimate of the linear attenuation coefficients of microcalcifications μ_{mc} and background tissue μ_b, and the film curve gradient c_1.

$$C_i^{mc} = y_i - y_B = c_0 c_1 (\log e) d_{mc}(\mu_{mc} - \mu_b) \qquad for\ y_i \in R, \tag{2.6}$$

From equation 2.6 it is clear that if logarithmic scaling is applied, microcalcification pixel contrast is approximately independent of breast thickness, exposure level and digitization. Using a fixed contrast threshold a_C, more consistency is expected in segmentation with a threshold T_c defined by

$$T_C = y_B + a_C. \tag{2.7}$$

Local signal dependent threshold

For thin objects the threshold value should be close to y_B. For thick objects this might not be a good choice. A low threshold value may result in overestimation of object size for thicker objects since blurring leads to higher intensity of pixels adjacent to the object region. The local signal dependent criterion, explained below, is expected to be insensitive for this. In this method the threshold level will be increased with the signal strength.

A small window of N pixels is placed in the center of the data field. In this window the mean intensity $\bar{y}$ is calculated as a measure for signal intensity.

$$\bar{y} = \frac{1}{N} \sum_{i=0}^{N-1} y_i \tag{2.8}$$

Typically, we use a 5×5 window in our experiments. In the segmentation process, a pixel is marked if its value exceeds a certain level T_s defined as

$$T_s = y_B + a_s(\bar{y} - y_B). \tag{2.9}$$

With a_s a constant with values between zero and one.

Iterative Method

To reduce the influence of noise on segmentation, we used an iterative segmentation approach based on a Markov random field. In a Markov Random Field (MRF) for each pixel a neighborhood is defined [18]. The MRF model is specified by giving the conditional probability distribution of a pixel label given its grey level and the labels of its neighbors. A Gaussian model is used for representing the fluctuation of grey levels due to noise:

$$\log P(x_i = l | (y_i, other\ labels)) \propto -\alpha(l) + \beta g(l) - (y_i - \mu_l)^2/(2\sigma_l^2), \tag{2.10}$$

where x_i is a label of pixel i which can take two values $l = 0, 1$ (background and foreground), and $\alpha(l)$ is an offset value. The iteration parameter β models the a priori likelihood of labels to occur close to each other where $g(l)$ is the number of neighbors with class l. The neighborhood that is used consists of the four nearest neighbors. Mean and standard deviation of the Gaussian model are estimated from background region B (μ_0 and σ_0) and from a 5×5 pixel window placed in the center of the data field (μ_1). We took σ_1 equal to σ_0 because screen-film system noise is expected to be relatively constant in the small data fields that are analyzed, since the average grey level within a field does not change much. Moreover, the microcalcification region is too small for a reliable estimation of σ_1. The values of $\alpha(l)$ and β are determined empirically.

In the segmentation process pixel labels x_i are iteratively updated by maximizing their probability according to equation 2.10 above. This approach is called estimation by Iterated Conditional Modes (ICM) as described in [19]. It should be noted that maximization of the expression in equation 2.10 requires an initial estimate of the labeling. This can be obtained by applying equation 2.10 with $g(l)$ set to zero. Next, $g(l)$ is updated before every new iteration step. A pixel is marked with the label that gives the highest probability. We applied 5 iteration steps since it appeared that the segmentation results become stable within 5 cycles.

2.2.3 Segmentation and background trend correction

The background structure of phantom recordings is relatively homogeneous compared to mammograms. To test segmentation in a more realistic situation we constructed images composed of phantom recordings with a mammographic background superimposed. From 5 clinical cases we took 10 mammograms (2 per case). A total of 391 data fields of 40×40

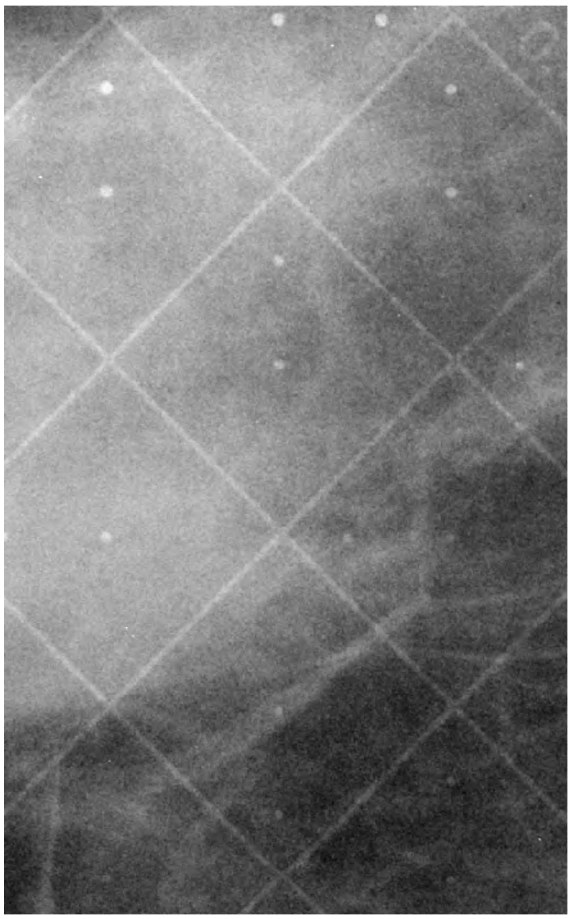


Figure 2.3: A part of a phantom recording superimposed on a part of a mammogram.

pixels containing background tissue were selected from these mammograms. Then, for every dot involved in the experiments a randomly selected mammographic data field was added to the 40×40 pixel data field containing the dot. Fig. 2.3 gives an impression of a phantom recording superimposed on a mammographic background. To perform an accurate segmentation of these dots in a mammographic structure a background trend correction is applied.

As before, we assume that a detection step has been carried out so that the positions of the dots are known. To define a background area a disc with a diameter of 1 *mm* is used. Microcalcifications will be covered by this disc. The center of the disc coincides with the center of the data field. The pixel values in the disc area R' are replaced by new pixel values interpolated from the surrounding background. Pixel y_i, with $i \in R'$ is replaced by y_i' according to the following weight function

$$y_i' = \frac{\sum_{k \in B} y_k / d_{ik}}{\sum_{k \in B} \frac{1}{d_{ik}}}, \tag{2.11}$$

where d_{ik} is the Euclidean distance between site i and site k, with k a pixel on the boundary

L of R'. The background image is low-pass filtered using a 9×9 uniform kernel and then subtracted from the original image. In the second stage, the segmentation is performed on the resulting image.

Instead of using a fixed disc to cover the object to be segmented, an initial segmentation could be used for creating a background image. However it turned out that low contrast dots can not be segmented in a consistent way, since background level and signal strength estimates are influenced by background structures. This may cause segmentation of high intensity background structures or severe underestimation of the objects. Therefore, a background trend correction based on an initial segmentation does not give convenient results for low contrast objects.

A contrast detail curve was determined visually for a CDMAM phantom with mammographic background superimposed. For this purpose a mammogram was selected and added to a phantom recording ($OD = 1.84$). Only the dots that were still visible were used in the experiments with the processed phantom images.

2.2.4 Contrast measurement

Contrast is defined by Equation 2.6 as the difference in pixel value of an object pixel and the mean background value. However, the effect of MTF is ignored in equation 2.6. Microcalcifications with equal object contrast will differ in radiographic contrast (i.e., the density difference between the imaged object and background) because of differences in object area.

Object contrast of a microcalcification is calculated by computing the mean of the contrast values for all pixels in the segmented region and by taking a correction term (A_0/A) for the MTF into account [20]

$$C^{mc} = (A_0/A)\frac{1}{N}\sum_{i=0}^{N-1} C_i^{mc}, \tag{2.12}$$

where N is the number of pixels in the segmented microcalcification region. A is the area of the object and A_0 is given by

$$A_0 = [\int\int |T(f)|^2 \, |MTF_s(f)|^2 \, df_x df_y]^{-1}, \tag{2.13}$$

where $T(f)$ is the Fourier spectrum of the object and $MTF_s(f)$ is the combined modulation transfer function of the screen/film system and digitizer. In first approximation contrast obtained by the equation above is independent of the exposure to the breast and breast thickness.

2.2.5 Evaluation methods

In order to compare performances of different segmentation methods we use two criteria representing accuracy and precision of the estimates: 1) the standard deviation $s(d)$ of estimated dot diameters with real diameter d, and 2) a measure D for the mean difference

between real diameter and estimated diameter

$$D = \frac{1}{N}\sum_i d_i - \hat{d}_i, \tag{2.14}$$

with N the total number of dots involved, $\hat{d}_i$ the estimated diameter of a dot with true diameter d_i. Small values for $s(d)$ indicate small variation between estimated diameters and a small D value indicates that the bias of the estimated diameters is small. A negative D value indicates a large fraction of overestimated dot diameters, where a positive D value indicates the opposite. It is noted that for the bias D a correction can be implemented if it is known. Optimization of segmentation parameters will be performed by minimization of the accuracy $s(d)$.

2.3 Results and discussion

Method	Parameters	*OD*=1.84		*OD*=1.21		*OD*=0.55	
		$\bar{s}$	D	$\bar{s}$	D	$\bar{s}$	D
LNDT	$a_n = 1.8$	0.13	-0.08	0.13	-0.09	0.14	-0.08
FCT	$a_C = 30$	0.11	-0.04	0.11	-0.07	0.11	-0.03
LSDT	$a_S = 0.5$	0.06	-0.01	0.06	-0.02	0.07	-0.02
MRF	α=0.1; β=1.5	0.05	-0.01	0.06	-0.02	0.06	-0.01

Table I: Accuracy and precision results for different segmentation methods and optical densities.

Phantom recordings were taken at three different exposure levels (optical density *OD* = 0.55, 1.21 and 1.84). In all the experiments in this section, computations are performed on 8 recordings for each exposure level. For each combination of diameter and thickness 16 dots can be retrieved from the 8 recordings since each phantom cell contains 2 dots.

In Fig. 2.4 results for recordings related to *OD* = 1.84 are shown, for respectively the local noise dependent threshold (a), the fixed contrast threshold (b), the signal dependent threshold (c) and the iterative method (d). The Figures correspond with optimized parameter settings according to the accuracy criterion. This is done by running the segmentation program for a large number of parameter values from a relevant interval for recordings with *OD* = 1.84. Then results with highest accuracy were selected. These optimized parameter values appeared to be good choices as well for recordings with lower optical density. The left-hand plots in Fig. 2.4 show the relation between mean estimated dot diameter and true dot diameter for different thicknesses given in μm. The mean estimated diameter was calculated over 16 estimated diameters, for dots originating from the same cell and retrieved from 8 recordings.

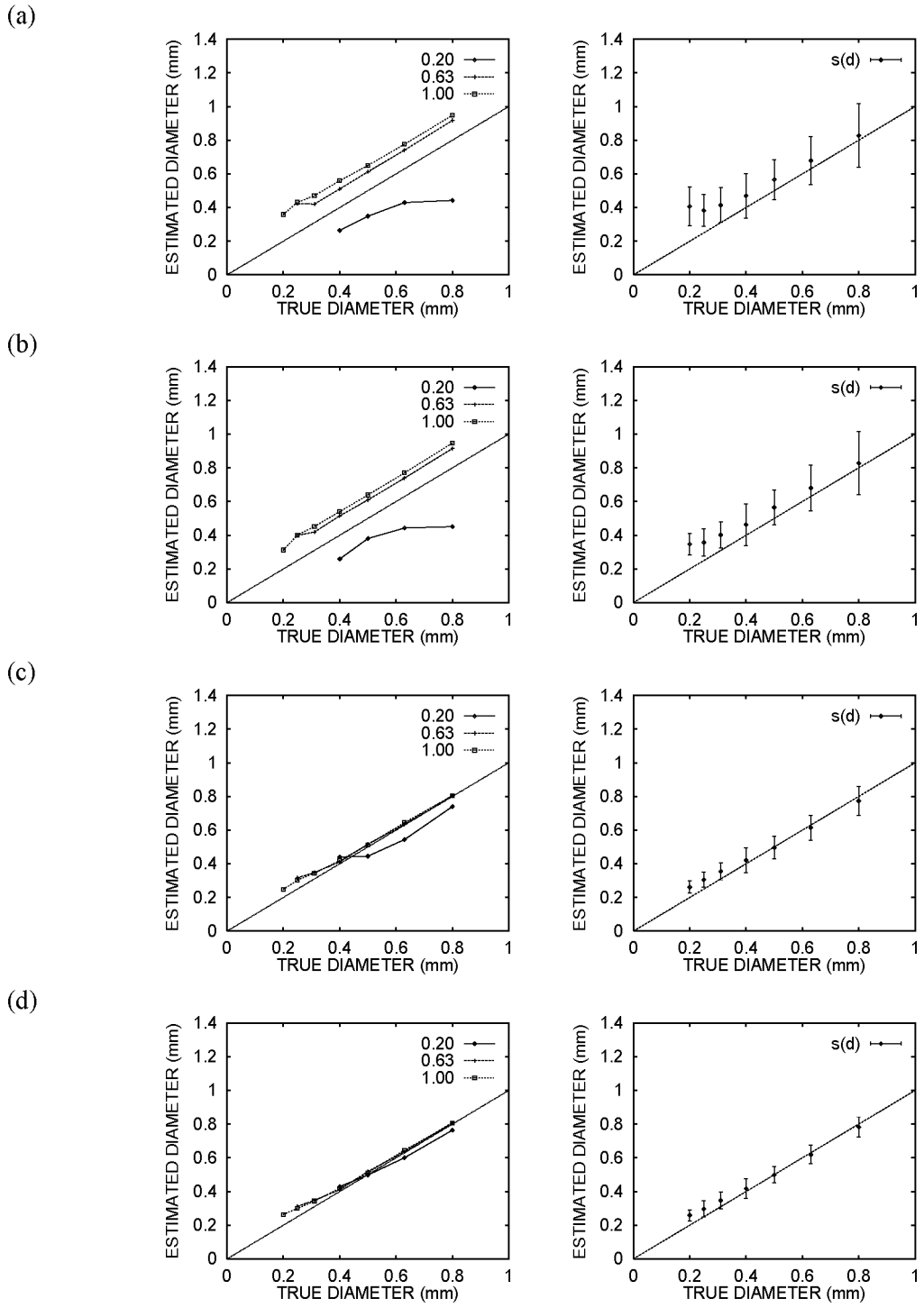


Figure 2.4: True diameter versus estimated diameter for three different thicknesses (left-hand plots; 20, 63 and 100μm). The right-hand plots show the accuracy $s(d)$ for all dots involved. Four methods are compared. (a) local noise dependent threshold; (b) fixed contrast threshold; (c) local signal dependent threshold; (d) iterative method.

Three different thicknesses are plotted including gold thickness = 20 μm, the lowest object contrast that was considered. The right-hand plots in Fig. 2.4 give the standard devi-

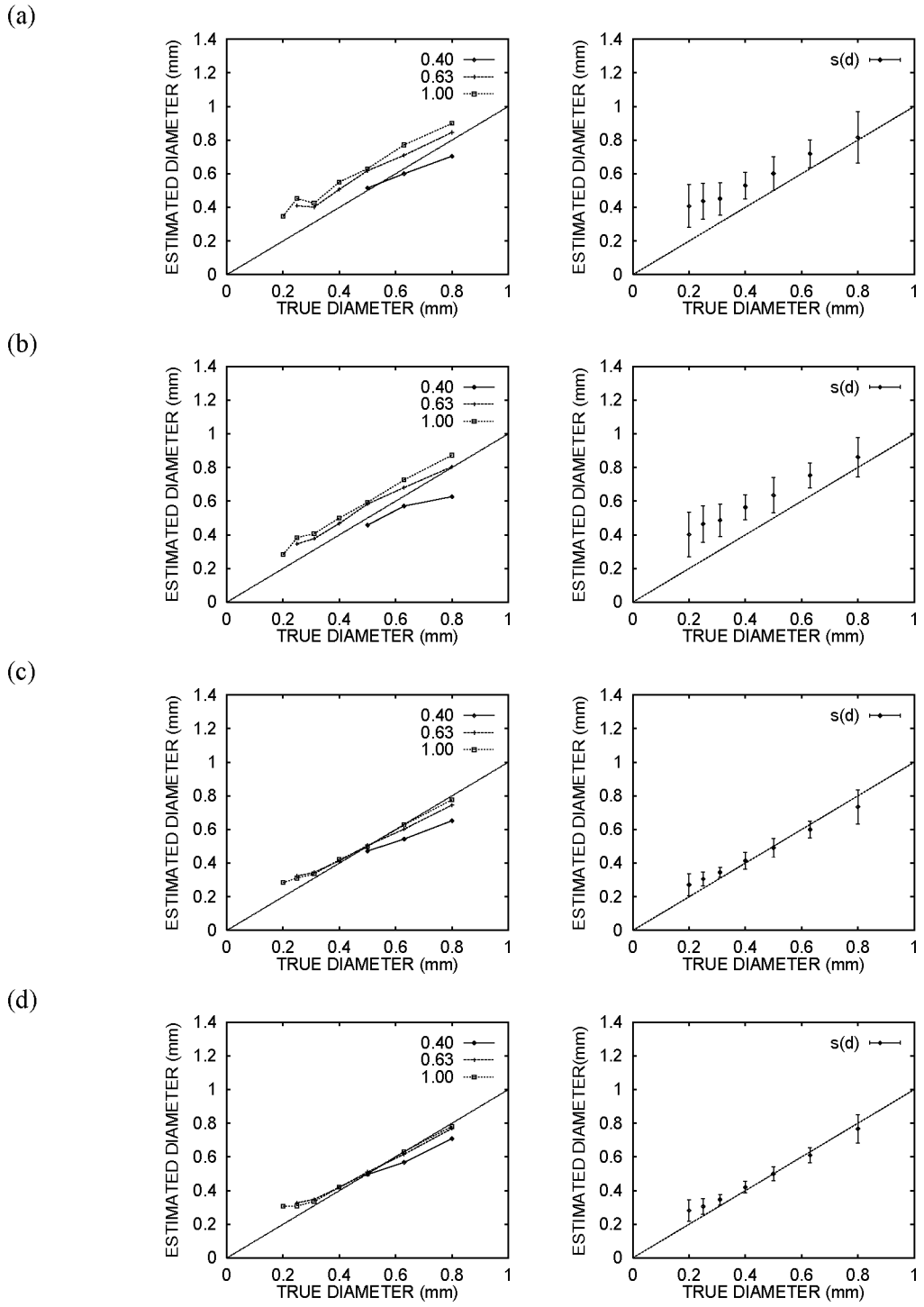


Figure 2.5: True diameter versus estimated diameter for three different thicknesses (left-hand plots; 20, 63 and 100μm). The right-hand plots show the accuracy $s(d)$ for all dots involved. Four methods applied on phantom images with mammographic background are compared. (a) local noise dependent threshold; (b) fixed contrast threshold; (c) local signal dependent threshold; (d) iterative method.

Method	Parameters	OD=1.84	
		$\bar{s}$	D
LNDT	a_n=2.2	0.11	-0.11
FCT	a_C=60	0.09	-0.07
LSDT	a_s=0.6	0.05	0.04
MRF	α=0.0; β=1.5	0.05	-0.01

Table II: Accuracy and precision results for different segmentation methods applied on images with mammographic background.

ation values $s(d)$ for all dots involved. The standard deviation $s(d)$ was calculated over all estimated dot diameters with true diameter d retrieved from 8 recordings.

From Fig. 2.4 it is clear that diameters of thicker dots are segmented larger in size than thinner dots when the local noise dependent threshold and the fixed contrast threshold are used. The difference in segmentation of thicker dots versus thinner dots causes large inaccuracies as shown in $s(d)$. In these methods the threshold has to be chosen close to the background level y_B in order to segment thin objects. For thicker objects a low threshold value results in overestimation of object size due to blurring, in the imaging system, which leads to higher intensity of pixels adjacent to the object region.

The local signal dependent threshold and the iterative MRF segmentation are signal strength dependent and are less sensitive for signal blurring. Therefore these methods give more consistent results. Furthermore, we found that the iterative method is less sensitive for inaccuracy in diameter estimation for thin dots caused by lower signal to noise ratios. In both methods, an overestimation of diameters is seen for dots in the range of 0.2 *mm* to 0.4 *mm*. This phenomenon is partly caused by the fact that the initial measure of signal strength is determined in a window that is rather large compared to the smallest dots. This means that a part of the background signal is measured and the estimation of signal strength will be too low in this case. Furthermore, intensity values in small dots are also underestimated because of signal blurring. One could use a smaller window in order to obtain more accurate signal strength estimates. However, it was found that a 3×3 window deteriorates the overall accuracy. A correction can be implemented to compensate for the overestimation of small dots.

Performances of different segmentation methods are given in Table I for recordings related to OD = 0.55, 1.21 and 1.84. In Table I the bias values D and mean standard deviation values $\bar{s}$ are given, where the latter were calculated as the mean value of standard deviation $s(d)$ over all d's. These results correspond with optimized parameter settings with respect to $\bar{s}$ (Table I). It appears that overall results are hardly affected by the exposure level. An effect of exposure level on segmentation accuracy was only observed for dots with very low contrast, close to the contrast detail threshold curve.

a)

b)

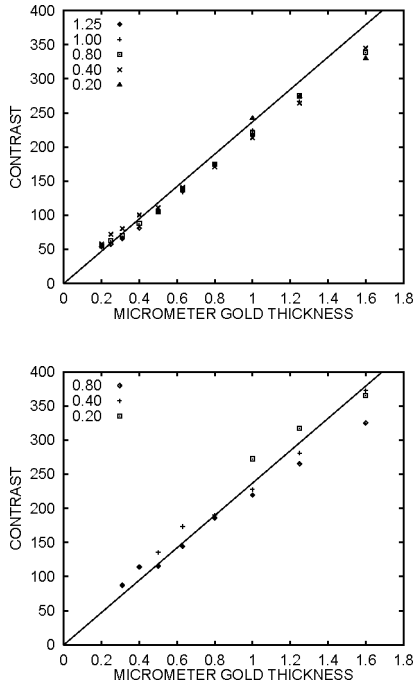


Figure 2.6: Contrast measurements corrected for MTF for different diameters given in *mm*. Gold thickness is given in μm. Calculated theoretical contrast is represented by the solid line. (a) homogeneous background; (b) mammographic background.

The local noise dependent criterion and the fixed contrast criterion give unsatisfying results. It appears that the iterative method based on a Markov random field and the local signal dependent criterion give much better results. The performances of both methods are comparable.

In Fig. 2.5 results related to the segmentation methods applied on phantom recordings ($OD = 1.84$) with mammographic background are shown. These results should not be compared directly to those in Fig. 2.4, computed on the unprocessed phantom images, because different contrast detail curves were applied in both cases for selecting the dots used in the experiments. In Table II the bias values D, the mean standard deviation values $\bar{s}$ and the optimized parameter values are given. Also for films with mammographic background, the iterative method and the local signal dependent criterion give best results and comparable performance. Except for the fixed contrast criterion, the optimized parameter values are similar to those in Table I for unprocessed images. The threshold parameter of the fixed contrast criterion is not defined relative to noise level or signal strength and becomes twice as large, because the noise level is roughly two times higher in the processed images.

Finally, Fig. 2.6 shows contrast measurements when corrected for the MTF. Corrected contrast is plotted against the object thickness of the gold dots. The iterative method is used for segmentation in both Figures. Fig. 2.6(a) corresponds to unprocessed phantom images while 2.6(b) corresponds to images with superimposed mammographic background.

For both cases recordings with optical density $OD = 1.84$ were used. A theoretical relation between object thickness and contrast was calculated using equation 2.6 and $\mu_{gold} = 190$ mm^{-1} [21], taken at 18 keV. A small deviation of the measurements from the theoretical value (Fig. 2.6) is probably due to beam hardening effects which are neglected in the mono-energetic approximation that we used. An increase of radiation energy in the perspex layer on the top of the phantom will decrease the gold attenuation factor and thereby object contrast. It is shown that radiographic contrast is approximately linear with object thickness which is in accordance with equation 2.6. As expected, for very low signal to noise ratios contrast estimation becomes inaccurate. The minimum object contrast needed for accurate contrast measurement, increases for dots smaller in size. This is explained by the fact that the signal to noise ratio decreases with decreasing object area, due to both increased influence of noise and signal blurring [20]. It is found that for diameters larger than 1.0 mm the effect of reduction of contrast by blurring is only small. For smaller dots, that are in the range of sizes that represent microcalcifications, the effect is significant (e.g. we found $A_0/A = 1.3$ for dot diameter $d = 0.4$ mm and $A_0/A = 2.0$ for dot diameter $d = 0.2$ mm). Apart from correcting for MTF it could also be considered to create an inverse filter in order to restore the object shape degradation caused by blurring [22]. In Fig. 2.6(b) a background trend correction is performed which limits dot diameters to 0.8 mm because of the 1 mm disc used for defining the background region. It appeared to be essential to use the corrected images for measuring contrast since a background trend hampers estimation of the mean background level. In approximation, the high frequency noise level in these images is two times higher. Therefore, roughly twice as much object thickness is needed here for obtaining reliable contrast estimates. Finally, in Fig. 2.6(b), contrast estimation for small dots ($d = 0.2$ mm) is less accurate than for larger dots. This due to noise which deteriorates segmentation performance for small dots.

2.4 Conclusion

Shape and contrast features of microcalcifications are important features in classification of malignant and benign microcalcification clusters and in differentiating between true positive and false positive detections. Therefore, accurate segmentation of microcalcifications is essential.

In this phantom study it could be shown that thresholding methods based on a local noise dependent threshold or a fixed contrast threshold give very poor results yielding size and contrast measurements that can hardly be used for classification purposes. By using a signal dependent threshold or the iterative method, based on a Markov random field model, much better results could be obtained. The segmentation performances of both methods are comparable and also in mammographic images these methods work well provided that a background trend correction has been carried out.

In this study contrast is defined in a way that influence of breast thickness, exposure level and digitization is eliminated. In measuring contrast a correction factor has to be applied since contrast of small objects is degraded by the MTF of the imaging chain.

Currently we are applying the results obtained in this study to improve our classification method of microcalcifications, using corrected size and contrast measurements.

2.5 Acknowledgments

The authors wish to thank the Dutch Cancer Society for supporting this work by grant KUN 95-897.

Bibliography

[1] R M Nishikawa, M L Giger, K Doi, C J Vyborny, and R A Schmidt. Computer-aided detection and diagnosis of masses and clustered microcalcifications from digital mammograms. In K W Bowyer and S M Astley, editors, *State of the art in digital mammographic image analysis*, volume 9 of *Series in machine perception and artificial intelligence*, pages 82–102. World Scientific, 1994.

[2] N Karssemeijer, J T Frieling, and J H Hendriks. Spatial resolution in digital mammography. *Invest radiol*, 28(5):413–9, May 1993.

[3] T Netsch. scale-space approach for the detection of clustered microcalcifications in digital mammograms. In K Doi, M L Giger, R M Nishikawa, and R A Schmidt, editors, *Digital Mammography*, pages 301–306. Elsevier, Amsterdam, 1996.

[4] J Parker, D R Dance, D H Davies, L J Yeoman, M J Michell, and S Humphreys. Classification of ductal carcinoma in-situ by image analysis of calcifications from mammograms. *Proc SPIE 1905*, pages 832–840, 1993.

[5] M Zhang and M L Giger. Automated detection of spiculated lesions and architectural distortions in digitized mammograms. *SPIE 2434*, pages 846–855, 1995.

[6] R N Strickland and H I Hahn. Wavelet transforms for detecting microcalcifications in mammograms. *IEEE Transactions on Medical Imaging*, 15(2):218–228, 1996.

[7] H P Chan, Lo SCB, L T Niklason, D M Ikeda, and K L Lam. Image compression in digital mammography: Effects on computerized detection of subtle microcalcifications. *Medical physics*, 23(8):1325–1336, 1996.

[8] F Levebvre, R Gilles, J Masselot, and E Kahn. Computer characterization of clustered microcalcifications in mammograms. In H U Lumke, M L Rhodes, C C Jaffe, and R Felix (Springer-Verlag Berlin 1991), editors, *Computer Assisted Radiology Proceedings of the International Symposium*, 1991.

[9] Y L Jiang, R M Nishikawa, D E Wolverton, C E Metz, M L Giger, R A Schmidt, C J Vyborny, and K Doi. Malignant and benign clustered microcalcifications: Automated feature analysis and classification. *Radiology*, 198:671–678, 1996.

[10] J Parker and D R Dance J Suckling. The hysteresis technique for detection of calcifications in digital mammograms,. In A G Gale, S M Astley, D R Dance, and A Y Cairns, editors, *Digital Mammography*, pages 41–50. Elsevier, Amsterdam, 1994.

[11] J Dengler, S Behrens, and J.F. Desaga. Segmentation of microcalcifications in mammograms. *IEEE Trans. on Medical Imaging*, 12(4):634–642, May 1993.

[12] I Brodie and A Gutcheck. Radiographic information theory and application to mammography,. *Med. Phys.*, 9:79–95, 1982.

[13] A D A Maidment, M Albert, E F Conant, and S A Feig. Three-dimensional calcification reconstruction from a limited number of views. *Proc. SPIE*, 2708:378–389, 1996.

[14] W J H Veldkamp and N Karssemeijer. Influence of segmentation on classification of microcalcifications in digital mammography. In *18th Annual international conference of the IEEE engineering in medicine and biology society*, 1996.

[15] N Karssemeijer and M A O Thijssen. Determination of contrast-detail curves of mammography systems by automated image analysis. In K Doi, M L Giger, R M Nishikawa, and R A Schmidt, editors, *Digital Mammography*, pages 155–160. Elsevier, Amsterdam, 1996.

[16] P K Sahoo, S Soltani, A K C Wong, and Y C Chen. A survey of thresholding techniques. *Computer vision, graphics and image processing*, 41:233–260, 1988.

[17] R M Nishikawa, M L Giger, K Doi, C J Vyborny, and R A Schmidt. Computer-aided detection of clustered microcalcifications: an improved method for grouping detected signals. *Med Phys*, 20(6):1661–1666, 1993.

[18] B D Ripley. Statistics, images, and pattern recognition. *The Canadian journal of statistics*, 14:83–111, 1986.

[19] J Besag. Statistics, on the statistical analysis of dirty pictures. *Journal of Statistical Society B.*, 48:259–302, 1986.

[20] G T Barnes and D P Chakraborty. Radiographic mottle and patient exposure in mammography. *Radiology*, 145:815–821, 1982.

[21] J H Hubbell. Photon mass attenuation and energy-absorption coefficients from 1 kev to 20 mev. *Int. J. Appl. Radiat. Isot.*, 33:1269–1290, 1982.

[22] R C Gonzalez and P Wintz. *Digital image processing*. Addison-Wesley Publishing Company, Inc., 4 edition, 1987.

Chapter 3

Normalization of Local Contrast in Mammograms[1]

Abstract

Equalizing image noise has been shown to be an important step in automatic detection of microcalcifications in digital mammograms. In this study, an accurate adaptive approach for noise equalization is presented and investigated. No additional information obtained from phantom recordings is involved in the method, which makes the approach robust and independent of film type and film development characteristics. Furthermore, it is possible to apply the method on direct digital mammograms as well. In this study, the adaptive approach is optimized by investigating a number of alternative approaches to estimate the image noise. The estimation of high frequency noise as a function of the grey scale is improved by a new technique for dividing the grey scale in sample intervals and by using a model for additive high frequency noise. It is shown that the adaptive noise equalization gives substantially better detection results than a fixed noise equalization. A large database of 245 digitized mammograms with 341 clusters was used for evaluation of the method.

3.1 Introduction

We are developing a diagnostic method for assisting radiologists in detection and interpretation of microcalcification clusters in mammograms. Microcalcifications may appear as an early sign of breast cancer and play an important role in diagnosing mammograms. In this study we focus on automated detection of microcalcification clusters. Histologically, it has been found that there are usually much more microcalcifications than are seen on a mammogram. It has been shown that it is not the resolution, but the noise which limits detection. Studies suggest that, due to high frequency noise, isolated spherical microcalci-

[1]This chapter is based on: W.J.H. Veldkamp, N. Karssemeijer, *Normalization of Local Contrast in Mammograms*, to appear in IEEE Transactions on Medical Imaging.

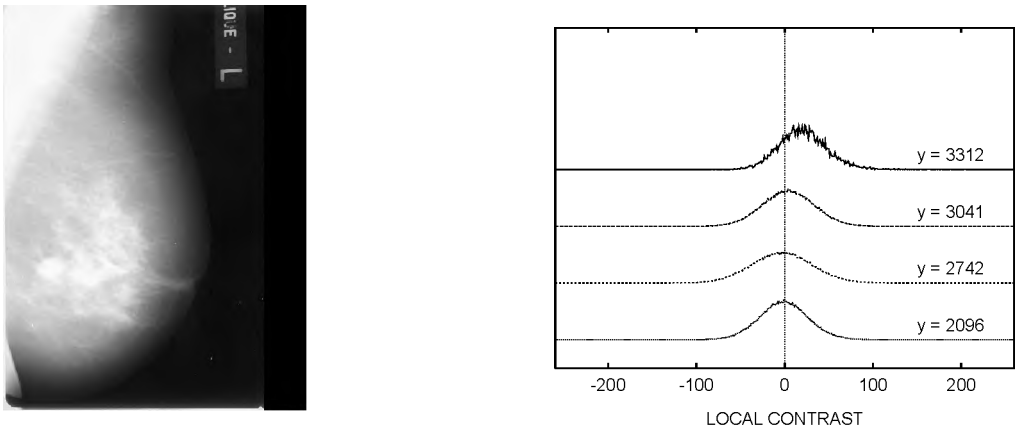
fications smaller than 130 μ can not be detected with film-screen mammography [1, 2]. In order to detect microcalcifications, therefore, reliable measurement of the image noise is of crucial importance.

Most microcalcification detection methods described in literature use some adaptive threshold that is locally determined. This is important for dealing with variation of the noise level across the image. Nishikawa et al. use a global grey level threshold in a first stage and a local adaptive thresholding technique in a second stage [3]. Chan et al. use local grey-level thresholding. The local threshold is variated with the standard deviation of the surrounding pixel values [4]. Kegelmeyer and Allmen analyze six different algorithms, including three algorithms based on grey level thresholding and three algorithms that use local contrast estimation [5]. Local contrast images are then globally or locally thresholded to find microcalcifications.

Adaptive thresholds that are determined locally in small image regions do not scale with the image noise only, but are influenced by image structures like lines and edges as well. Therefore, in regions with a lot of image structure, thresholds will not be adjusted optimally to the high frequency noise level and detection performance may deteriorate. In previous work it has been shown that the use of adaptive thresholds that are locally determined can be avoided by taking the signal dependency of noise into account [6]. A statistical microcalcification detection method was developed that was combined with a preprocessing step, in which images were rescaled to equalize image noise. This preprocessing step transforms the input image in which noise depends strongly on the signal level to an image with a homogeneous noise level. In order to normalize the input image, high frequency noise is determined as a function of the grey level from the image itself, and from this information a correction is performed. Such an adaptive approach does not rely on the stability of the image formation process, and takes tissue inhomogeneity into account as an additional noise component. The importance of modeling signal dependency of the noise for detection of microcalcification clusters is confirmed in other studies. Netsch et al. used a method based on the Laplacian scale-space representation of the mammogram and applied a noise equalization [7]. Maitournam et al. used splines to model the trend in each mammogram. The trend image was then subtracted from the original image. The noisy result image was thresholded after a variance equalization was done [8]. Brown et al. used a wavelet approach followed by a second step of feature based classification. They emphasized that an accurate noise equalization is essential for their detection method [9]. Other approaches are based on adaptive noise suppression and other hybrid wavelet adaptive enhancement methods [10, 11, 12, 13, 14]. A disadvantage of these latter methods is that they suppress the noise (or enhance the signal) locally without taking the signal dependency of the noise into account.

In the method described previously, features representing the image data are calculated from the rescaled image. For feature extraction it is the uncertainty in feature space which

(a)



(b)

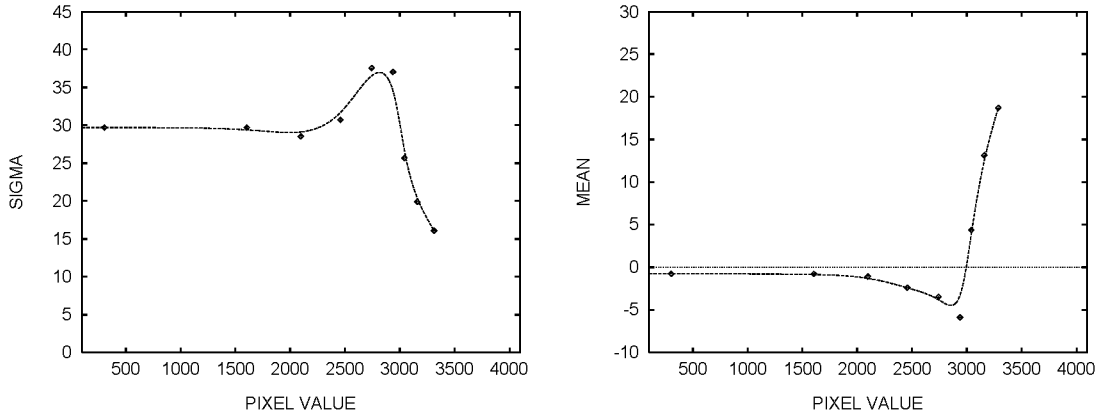


Figure 3.1: Noise characteristics are estimated for each mammogram separately. Figure (a) shows histograms of local contrast calculated from a mammogram (digitized at 12 bits) for four different bins. The vertical axis represents the frequency of occurrence. Also the central pixel values of the bins are shown in the figure. We used a 9×9 window for calculating local contrast. Figure (b) shows the continuous functions $\sigma_c(y)$ and $\mu_c(y)$ obtained by interpolation of the estimates $\sigma_c(k)$ and $\mu_c(k)$ with $k = 1,2,..., 8$.

is relevant. Noise should therefore be related to the standard error of feature values. Instead of first converting the image itself to equalize the noise, it is also possible to normalize the features directly. The latter approach is chosen in this work. In the detection scheme, dependency on grey level y is removed from feature space by rescaling local contrast features using the standard deviation of local contrast $\sigma_c(y)$ as a measure for high frequency noise. Local contrast c_i at site i is usually defined by convolution with a filter function. We use a relatively simple filter which turns out to work well in comparison to many others [5, 15]:

$$c_i = y_i - \frac{1}{N} \sum_{j \in \partial_i} y_j, \tag{3.1}$$

where ∂_i represents a neighborhood or window at i of size N. For obtaining $\sigma_c(y)$ the grey-scale is divided in non-overlapping but adjacent bins (or bands) numbered $k = 1, 2, ..., K$.

After computing local contrast, the probability density function $f(c|k)$ can be estimated by normalizing the histograms of c determined within each bin k. The local contrast is calculated according to equation 3.1. Fig. 3.1(a) shows an example of local contrast distributions calculated from a 12-bits digitized mammogram for four different bins. We used a 9×9 window for ∂_i ($N = 81$). The corresponding histograms and central pixel values are shown in the figure. For each bin k the standard deviation $\sigma_c(k)$ of local contrast c can be calculated from $f(c|k)$, with $k = 1, 2,..., 8$. Fig. 3.1(b) shows the continuous function $\sigma_c(y)$ obtained by interpolation of the estimates $\sigma_c(k)$. Note that an extra point, at the minimum pixel value in the breast region, is inserted to guide the interpolation along the lowest pixel values. The corresponding σ_c value is taken identical to the estimate of σ_c in the corresponding bin that covers the lowest pixel values. It is clear from these figures that $\sigma_c(y)$ varies strongly with the grey level. Fig. 3.1 (b) also shows the mean of local contrast μ_c as a function of the grey level. Although from symmetry considerations μ_c would be expected to be zero, in higher regions of the grey scale μ_c appears to increase whereas in in lower grey scale regions the opposite effect occurs. In previous work this effect was mentioned but ignored. Taking the signal dependency of both σ_c and μ_c into consideration, it is possible to normalize local contrast c_i by:

$$c_i' = (c_i - \mu_c(y_i))/\sigma_c(y_i), \tag{3.2}$$

where c_i' represents normalized local contrast at site i.

In earlier work, described in [6], some additional information obtained from phantom measurement was used in normalization for image noise. A disadvantage of using such a model is that dependency on, for instance, film type and film development characteristics is introduced. The range of grey levels per bin or the bin width was chosen to increase exponentially with the grey level to obtain a more uniform distribution of the sites over the bins, where it should be remarked that the grey values in the images used were linear with intensity. After having obtained estimates $\sigma_c(k)$ for a number of bins k, the continuous function $\sigma_c(y)$ was estimated by interpolation using a polynomial fit where μ_c was assumed to be zero and thereby considered to be independent for signal intensity. To guide the interpolation, at the maximum grey level fixed values of σ_c and the slope of the curve had to be imposed, because in higher bins the number of pixels seemed often to small to estimate the noise level in a reliable way. These fixed values were obtained from phantom measurements.

In [7] a different implementation of the technique is presented by Netsch et al. that avoids using a model. The authors use overlapping bins, centered around each pixel value in the image, where each bin is build up from an equal number of pixels. In this way noise estimates can be obtained for every pixel value in the mammogram without any interpolation. Since for each pixel value a noise estimate is obtained, use of additional information from phantom measurement can be avoided. Using this idea we experimented with alternative approaches for normalizing image contrast. Interestingly, we found that relatively small

changes in the algorithm have substantial influence on the detection performance. Recently ([16]) we reported on an improved method for noise equalization which gave superior detection results compared to results obtained in earlier work where phantom information was involved in the process of normalization for image noise. We compared this new approach with our earlier method in [6] where phantom information was used.

In this paper we investigate a number of approaches for normalization of local contrast in detail without using phantom information. Among these is the improved method described briefly in [16]. We experimented with small bins where interpolation was omitted and larger bins consisting of a variable number of pixels where a B-spline interpolation was used. Also, the effect of the size of window ∂_i for calculating local contrast in equation 3.1, used for obtaining high frequency noise as a function of the grey level, on the detection performance was investigated. More precisely, it was examined how well the measure for noise estimation should correspond to the features used in the detection algorithm. As already discussed, histograms of local contrast are asymmetrical around zero in higher and lower bins, which causes non-zero mean values for local contrast and may hamper accurate noise estimation. It is investigated whether or not normalization of local contrast can be optimized by methods that involve correction for asymmetry. Finally, the adaptive approach for noise equalization is compared to using a fixed scale conversion derived from a phantom recording.

The method for detection is based on Bayesian techniques. A Markov random field model is used to model spatial relation between the labeled pixels in an iterative process. Three different features for representing the image data are used for detection: the local contrast at two different spatial resolutions and the output of a line/edge detector.

For this study a database was constructed by selecting all cases with microcalcification clusters from a four years period of breast cancer screening in Nijmegen. This data set consists of 245 mammograms from 125 women (of which 114 underwent biopsy) with 341 clusters. Labeling of the true clusters was done according to the radiological screening reports which contained schematic drawings of the location of true clusters. All clusters cases were selected for additional work up. The mammograms were digitized at 0.05 *mm* per pixel linear with optical density using a Lumisys 85 digitizer and were averaged down to 0.1 *mm* per pixel. To train the detection method, a different set of 25 mammograms was used. These training images where digitized at a 100 micron resolution using a 12 bits CCD Camera (Eikonix 1412). It should be noticed that the detection program is trained on images from a CCD camera, yielding pixel values linear with intensity, whereas the test images are logarithmically scaled to obtain values that are linear with optical density. However, as a result of normalizing the local contrast features for signal dependent noise, no extra conversion had to be applied. FROC curves were used for evaluation of the detection performance.

Finally we want to comment on the impact of image resolution in this study. For in-

stance, it should be noted that the filters used in the proposed methods and specifically the corresponding window sizes, are scaled at images of 0.1 *mm* per pixel.

3.2 Recognition of microcalcifications

The method that is used for detection of microcalcifications in digital mammograms is based on the use of Bayesian techniques and application of a Markov random field model, where the latter models the fact that microcalcifications occur in clusters. Starting from an initial segmentation the labeling is optimized by applying an iterative rule for updating pixel labels. The image data is represented by filtered versions of the original mammogram,which are thought to be important in distinguishing microcalcifications from other structures. Extraction of these features is described briefly in the next subsection. A principal advantage of the approach is that all information available, i.e. the image data, the current labeling and prior beliefs, is exploited simultaneously.

3.2.1 Feature extraction

The detection scheme uses three different features for representing the image data: the output of a line/edge detector and the local contrast at two different spatial resolutions.

The line/edge feature is calculated from the local probability density function of gradient directions, which is estimated by applying the Sobel operator as described in detail in [6]. This feature is meant for detecting structures which are linear within a given window size, as opposed to the blob-like microcalcifications. The use of such a feature is necessary because both thin lines and regions near strong gradients may easily give rise to false positive detections, the latter because of rippling of the local contrast at sharp boundaries.

Local contrast c_i at site i is defined by equation 3.1. In this work a 9×9 window is taken for ∂_i. The lower resolution local contrast is derived from normalized local contrast c_i' by smoothing the normalized local contrast image using a 3×3 uniform filter kernel.

3.2.2 Statistical model

During the detection process pixel labels x_i are iteratively updated by maximizing their probability, given the image data in a small neighborhood $y_{\delta i}$ of site i and given the current estimate of the rest of the labeling $\hat{X}_{S\setminus i}$:

$$x_i' = \max_l [p(x_i = l | y_{\delta i}, \hat{X}_{S\setminus i})], \tag{3.3}$$

where $l = 1,2,3,4$ represents four pixel classes: background, microcalcifications, lines/edge and film emulsion errors. The image data $y_{\delta i}$ is represented by the three local image features mentioned in the preceding subsection. The probability to be maximized can be written as

$$p(x_i = l | y_{\delta i}, \hat{X}_{S\setminus i}) \propto f(\Theta_i | x_i = l, \hat{X}_{S\setminus i})\, p(x_i = l | \hat{X}_{S\setminus i}), \tag{3.4}$$

where Θ_i is a vector denoting the values of the three features at a particular site. The *a priori* probability $p(x_i|\hat{X}_{S\setminus i})$ of the labels represents the Markov random field as described extensively in [6] and models spatial relations. For instance, a pixel is more likely to be part of a calcification if there are other calcifications in the neighborhood.

3.3 Optimizing local contrast normalization

As explained in the introduction, detection of microcalcifications primarily concerns separating image signals due to microcalcifications from those due to high frequency noise. An important step in the detection scheme, therefore, is a normalization of local contrast. This chapter describes methods which are intended to optimize the normalization of local contrast. We present two variants for dividing the grey scale into bins and will analyze their effect on the detection performance. Furthermore, approaches are described for handling asymmetry in histograms of local contrast.

3.3.1 Determination of bins

To avoid dependency of the method on knowledge of the image acquisition process, we experimented with two different approaches for determination of bins k needed for obtaining samples of $\sigma_c(k)$ and $\mu_c(k)$. First we used non-overlapping bins, consisting of a small number of pixels, without applying interpolation. Secondly, we used bins consisting of a variable number of pixels, and a B-spline interpolation to obtain a smooth function $\sigma_c(y)$.

Using a small number of pixels per bin

In this approach each bin consists of a predefined number of pixels M. As a consequence the width (i.e. the interval width) of the bins depends on the histogram of pixel values in the breast region. Bin width will be larger in parts of the grey scale which correspond with only few pixels whereas bin width we be smaller in parts with a larger amount of pixels. Values of M are chosen small, therefore interpolation between noise estimates can be omitted. This approach is based on the one suggested by Netsch et al. [7], who use overlapping bins however. Furthermore, their detection method uses Laplacian scale space theory and is thereby different from the one used in this study.

Using a variable number of pixels per bin

We found that it is difficult to obtain accurate noise estimates in the brightest region of the mammogram. This is explained by the fact that the standard deviation σ_c of local contrast decreases rapidly with increasing intensity in these brightest regions. This effect is hard to measure because of the relatively small number of pixels these regions often consist of. However, accurate noise estimation in the brightest region of a mammogram is important

because microcalcifications often appear in dense areas of the mammogram. As a consequence of improper noise estimates in dense regions, the detection scheme appeared to miss true clusters or generated relatively many false positive clusters.

To obtain more accurate estimates of σ_c in the brightest regions we used a variable number of pixels per bin M. The value M is chosen small for the highest brightness values because these regions consist of relatively few pixels. At lower brightness $\sigma_c(y)$ is more constant and M is chosen larger. It is remarked that we define bins starting from the maximum pixel value in the mammogram. In that way we avoid that a relatively small number of pixels remains in the brightest region, which might be insufficient for estimating noise accurately.

The first bin, covering the highest pixel values that occur in the image, consists of M_{min} pixels. The number of pixels $M(y_k)$ in each next bin k is determined by the fraction of brightest pixels in the mammogram $Z(y_k)$ which is defined as:

$$Z(y_k) = \int_{y_k}^{\infty} f(y)dy, \tag{3.5}$$

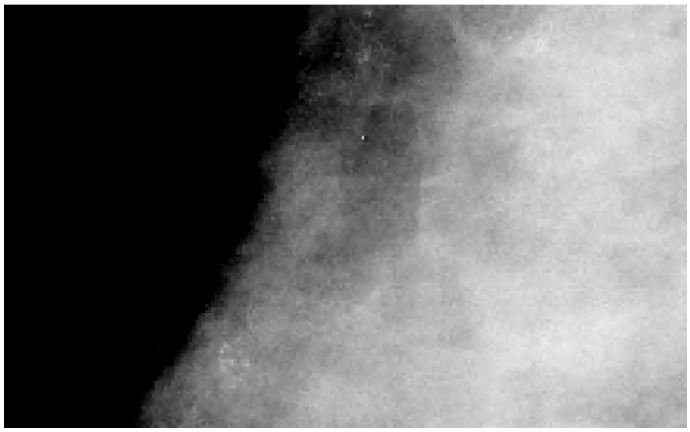
with y_k the highest pixel in bin k which is not involved in the preceding bins yet and $f(y)$ the probability distribution function of pixel values y in the image. The following relation describes the number of pixels per bin M, which is chosen to decrease linear for decreasing values of Z if $Z \leq Z_T$:

$$M(y_k) = \begin{cases} M_{max} & \text{if } Z(y_k) > Z_T \\ M_{min} + \frac{Z(y_k)}{Z_T}(M_{max} - M_{min}) & \text{if } Z(y_k) \leq Z_T, \end{cases} \tag{3.6}$$

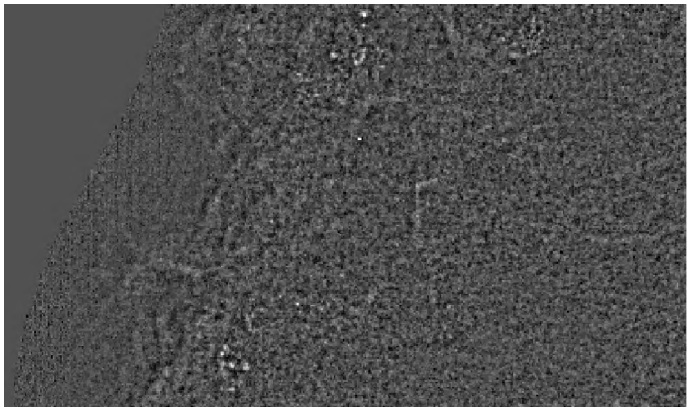
with Z_T a predefined fraction of brightest pixels in the mammogram (typically, $Z_T = 0.03$). The parameter M_{max}, is a predefined fraction of the number of pixels in the breast region of the mammogram.

For performing an interpolation between noise estimates, we used a B-spline interpolation as described in [17]. It appeared that in particular in the first bin, which is chosen rather small, noise estimation may not always be reliable. For instance, this the case when this bin covers part of a microcalcification cluster. High frequency noise will be overestimated in that particular case. To avoid this, we impose that the noise level in the first bin is equal or lower than the noise level in the second bin. If not, the noise estimate in the first bin is omitted and the interpolation curve is linearly extrapolated. In Fig. 3.2(a) shows part of a mammogram containing three clusters. Local contrast is shown in Fig. 3.2(b). The clusters occur in regions of different intensity levels. Cluster number two, appearing in a brighter region than cluster number one and three, is missed by the detection scheme when applying the method with a fixed number of pixels per bin. When using a variable number of pixels per bin the cluster in the brighter region is detected. Fig. 3.2(c) shows the corresponding detection output for the latter situation of a variable number of pixels per bin.

(a)



(b)



(c)

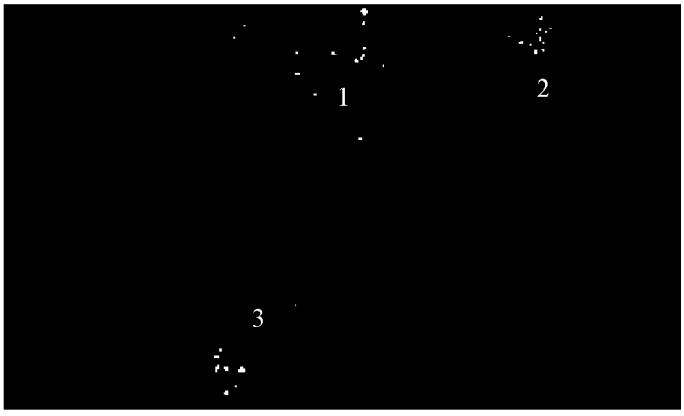


Figure 3.2: Part of a mammogram containing clusters. Three specific clusters are marked that occur in regions of different intensity levels. Cluster number two, appearing in a brighter region than cluster number one and three, is missed by the detection scheme when applying the method with a fixed small number of pixels per bin. When using a variable number of pixels per bin the cluster in the brighter region is detected as well.

3.3.2 Asymmetry in histograms of local contrast

In the introduction we explained how we normalize local contrast to remove signal dependency of the noise (equation 3.2). In this section we investigate the fact that in bins covering

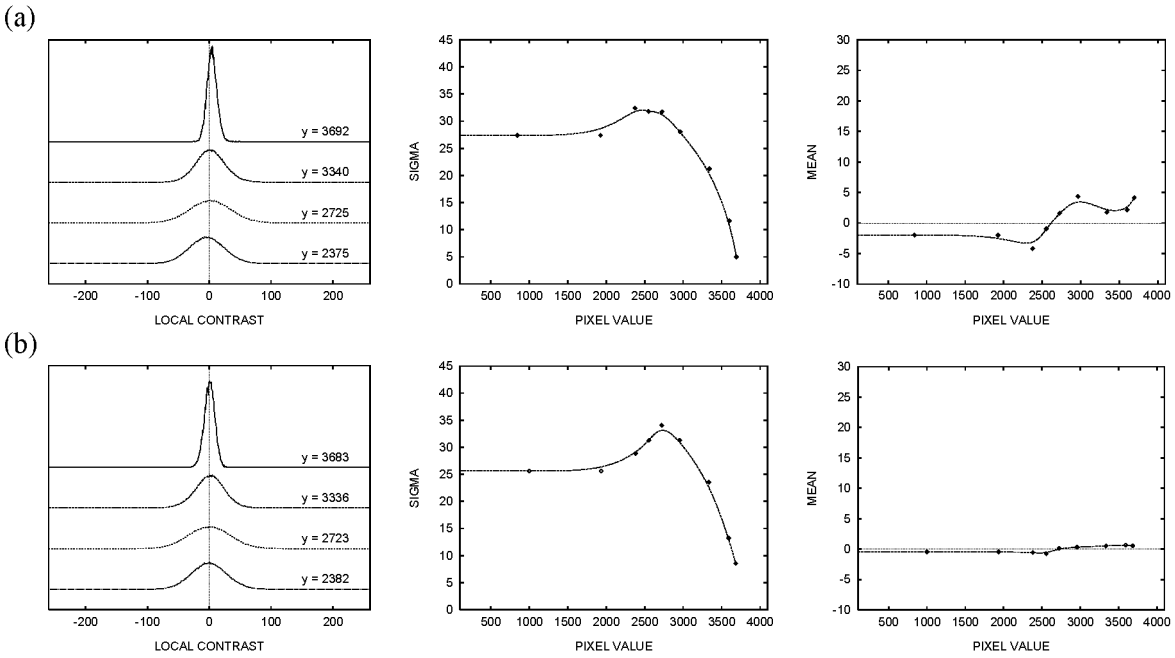


Figure 3.3: Asymmetry in histograms of local contrast calculated from a mammogram (digitized at 12 bits) for four different bins (left figures). Also the central pixel values of the bins are shown in the figure. The middle figures depict the standard deviation of local contrast as a function of the grey level whereas the figures at right depict mean as a function of the grey level. Fig. (a) corresponds to the method wherein μ_c is determined as a function of the grey level and Fig. (b) corresponds to bins wherein asymmetry is reduced.

the highest and lowest pixel values, histograms of local contrast are not symmetric around zero, particularly if the number of pixels per bin is chosen small. This is expected to deteriorate noise estimation.

The effect of asymmetry can be explained as follows. The brightest pixels in a mammogram will often have a positive mean local value, because statistically these pixels have a high probability of being surrounded by darker pixels. This is especially true for pixels at small local peaks of the intensity distribution. In lower bins the opposite effect occurs. However, when using larger bins in dark areas of the film asymmetry does not play a significant role.

It was investigated whether or not detection performance can be improved by correction for asymmetry. Different ways of dealing with asymmetric local contrast histograms, have been investigated. Apart from neglecting the effect by setting μ_c to zero, two correction schemes have been implemented. In the first method both μ_c and σ_c are determined as a function of the grey level for performing local contrast normalization. Secondly, a method is presented that reduces asymmetry in histograms by changing the sampling scheme.

Fig. 3.3 shows asymmetry in histograms of local contrast calculated from a mammogram (digitized at 12 bits) for four different bins (left figures). Also the central pixel values of the bins are shown in the figure. The middle figures depict the standard deviation of local

contrast as a function of the grey level whereas the figures at right depict mean as a function of the grey level. Fig. 3.3(a) corresponds to the method wherein μ_c is determined as a function of the grey level and Fig. 3.3(b) corresponds to bins wherein asymmetry is reduced.

Assuming μ_c to be zero

When assuming the effect of asymmetry to be neglectable, we can calculate the standard deviation of local contrast σ_c by setting the mean of local contrast μ_c to be zero in equation 3.2.

It should be noted that we are only interested in positive local contrast values since in the detection scheme negative local contrast values are associated with background tissue. Instead of using the entire histogram we can also consider the positive part of the histogram as the right-hand side of a symmetric distribution around zero. We determine the standard deviation belonging to this virtual symmetric distribution from the positive local contrast values only and do this for each bin.

Assuming μ_c to be grey level dependent

In this method we take the mean value μ_c as a function of the grey level into account. This can be achieved by determining $\mu_c(y)$ from the samples $\mu_c(k)$. This is achieved analogous to the determination of $\sigma_c(y)$ as described in the introduction and in section 3.3.1. Netsch et al. [7] use $\mu_c(y)$ to scale the detection threshold in his microcalcification detection scheme. We perform normalization according to equation 3.2, where the standard deviation $\sigma_c(k)$ is determined around $\mu_c(k)$. Fig.3.3 (a) gives an impression of the method. Normalization is performed according to equation 3.2.

Reducing asymmetry

In this section we make the assumption that a relation between a noisy pixel values y_i and pixel values without additional high frequency noise component, u_i can be written as [18]:

$$y_i = u_i + \eta_i, \tag{3.7}$$

where η_i is the high frequency noise contribution to u_i and where we consider c_i as a measure for η_i. If for each u_i the high frequency noise contribution c_i is known, and the grey scale is divided in bins composed from the pixel values u_i, then for each bin k the measure for high frequency noise $\sigma_c(k)$ can be determined. Since the noise processes are symmetric the mean of local contrast $\mu_c(k)$ is expected to be zero for each bin k.

In the previous sections, c_i was considered as the high frequency noise contribution to y_i instead of u_i. So a small error was made in linking the local contrast values to the corresponding grey levels. According to relation 3.7, selection of appropriate bins for each

pixel i should be based on the value of u_i, which can be estimated by smoothing of y_i. Using a linear filter denoted by:

$$v_i = \sum_{j \in \partial_i} a_j y_j, \tag{3.8}$$

with ∂_i a suitable chosen neighborhood at i of size N, and a_j a filter weight, substitution of relation 3.7 in 3.8 gives:

$$v_i = \sum_{j \in \partial_i} a_j u_j + \bar{\eta}_i, \tag{3.9}$$

where $\bar{\eta}_i$ is the weighted average of η_i. When, for instance, equal filter weights a_j are used and η_i is white noise with zero mean and standard deviation σ_η^2 then it follows that the variance $\bar{\sigma}_\eta^2 = \sigma_\eta^2 / N$. This means that the noise power is reduced by a factor N. If the noiseless image u_i is approximately constant over the window ∂_i and $\bar{\eta}_i$ is neglected, we can use v_i as a measure for u_i. In this study we used a Gaussian spatial filter for obtaining the pixel values v_i as an estimate for u_i. The filter size was determined experimentally (σ=12). In Fig. 3.3(b) it is shown that asymmetry is strongly reduced by this approach.

3.4 Performance evaluation

For evaluation of the detection performance the numbers of true and false positive clusters were determined for each mammogram, while the sensitivity of detection was varied. A FROC (free response operating characteristics) curve can be constructed by plotting the true positive fraction as a function of the mean number of false positive clusters per image. Labeling of the true clusters was done according to the radiologic screening reports which contained schematic drawings of the location of true clusters. For counting the number of true positives, a cluster was regarded as detected if two or more calcifications were found in the marked area covering the cluster. No verification of the detection of individual calcifications was performed. With respect to the false positives, a cluster was counted if a closed area was found in which two or more calcifications occurred. A closed area was determined by defining a disc of 1 *cm* in diameter, centered at each detected spot. A group of discs that touch or overlap forms a closed area. Clusters of large benign calcified cysts, benign fat necrosis calcifications and benign vascular calcifications are not regarded as true clusters in this study, and counted as false positives if detected.

3.5 Results

A database was constructed by selecting all cases with microcalcification clusters from a four years period of breast cancer screening in Nijmegen. The dataset used consists of

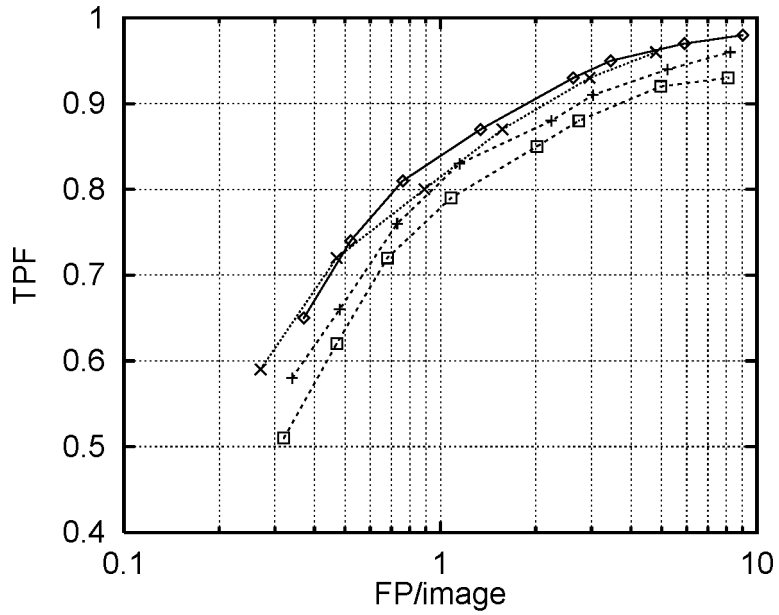


Figure 3.4: FROC curves showing the influence of adjustments in signal dependent noise correction on detection of microcalcification clusters, obtained on a dataset of 245 mammograms. Results related to noise estimation in relatively small bins (□) could be improved by using a larger and variable number of pixels per bin and B-spline interpolation (reference curve +). In both methods the entire histogram for estimating $\sigma_c(k)$ in each bin k is used and $\mu_c(k)$ is assumed to be zero. Normalization taking both μ_c and σ_c into account (×) improves results. Only the right hand side of the histogram with regard to the mean value $\mu_c(k)$ is used for determining $\sigma_c(k)$. Finally, the method of reducing asymmetry (◇) gives most convenient results.

245 mammograms with 341 clusters. The mammograms were digitized at 50 micron per pixel (12 bits pixels) with a Lumisys 85 digitizer and averaged down to 0.1 *mm* per pixel. The program was trained on a set of 25 mammograms with 60 microcalcification clusters different from the test set. These training images where digitized at a 100 micron resolution using a 12 bits CCD Camera (Eikonix 1412). As a result of normalizing the local contrast features for signal dependent noise, no extra conversion had to be applied.

Initially, we assumed the mean of local contrast μ_c to be zero and used the entire histogram in each bin k to determine $\sigma_c(k)$. In Fig. 3.4, the two lowest curves represent two different approaches for defining bins. The lowest curve was obtained when we choose the number of pixels per bin M to be small while interpolation between the noise estimates was omitted. We found that within a certain range, detection results were not very sensitive for variation of M. In the experiments we took $M = 5 \cdot 10^4$ pixels which appeared to give good detection results. An improvement is shown in detection performance when using a larger variable number of pixels per bin and applying a B-spline interpolation as described in subsection 3.3.1. The maximum number of pixels per bin M_{max}, is a predefined fraction of the number of pixels in the breast region and was in the order of $2 \cdot 10^5$ pixels. It was found experimentally that results deteriorate when using positive local contrast values only

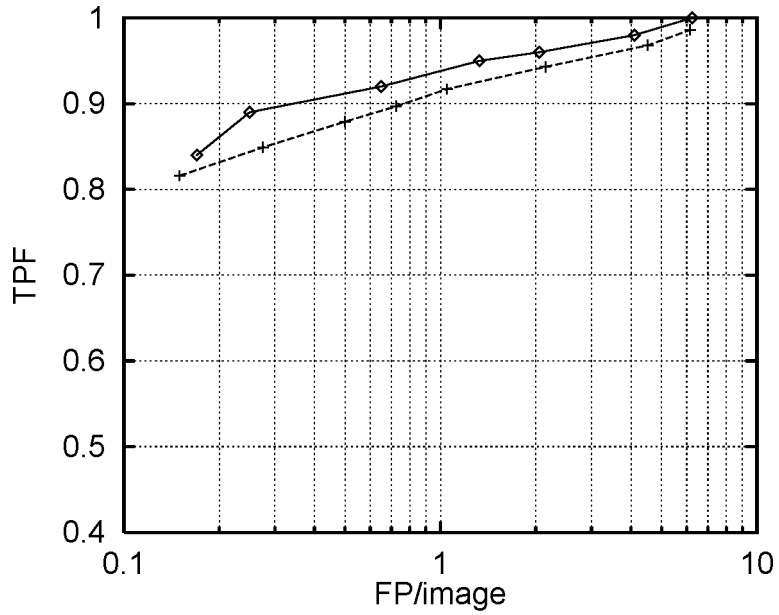


Figure 3.5: The method of reducing asymmetry (◇) shows an improvement compared to the phantom based method (+). Both methods are applied at a public data base of 40 images. Both the phantom based method an the data base are described in [6].

for determination of $\sigma_c(k)$ in each bin k compared to using the entire histogram as was the case in Fig. 3.4. In the experiments described next, we used large bins consisting of a variable number of pixels and B-spline interpolation. Therefore the second lowest FROC curve in the figure will be used as a reference for the following experiments, which are meant to overcome asymmetry in local contrast histograms.

In Fig. 3.4 also results of the method described in section 3.3.2 are given, where μ_c is considered as a function of the grey level. The mean of local contrast as a function of the grey level, $\mu_c(y)$, is determined from the samples $\mu_c(k)$. This is achieved analogous to the determination of $\sigma_c(y)$ by using a variable number of pixels per bin and B-spline interpolation. It was found that using this method gave improved results compared to assuming $\mu_c(k)$ to be zero as was the case for the reference curve. In each bin k only contrast values in the histogram higher than $\mu_c(k)$ were used for determination of $\sigma_c(k)$, which appeared to give more accurate results than when using the entire histogram for determination of $\sigma_c(k)$.

Fig. 3.4 finally gives FROC results related to the method of reducing asymmetry, as described in 3.3.2. It appears that detection results were improved compared to the reference curve. For instance, at a true positive fraction (TPF) of 0.91, the number of false positives per image (FP/image) decreases from 3 to 2. In addition, Fig. 3.5 presents the method of reducing asymmetry compared to the phantom based results from [6], where both methods are applied at a public data base of 40 images.

Fig. 3.6 demonstrates the effect of using a fixed scale conversion obtained from a phantom recording versus using an adaptive approach. It is shown that using a variable number of pixels per bin wherein asymmetry is reduced gives superior results.

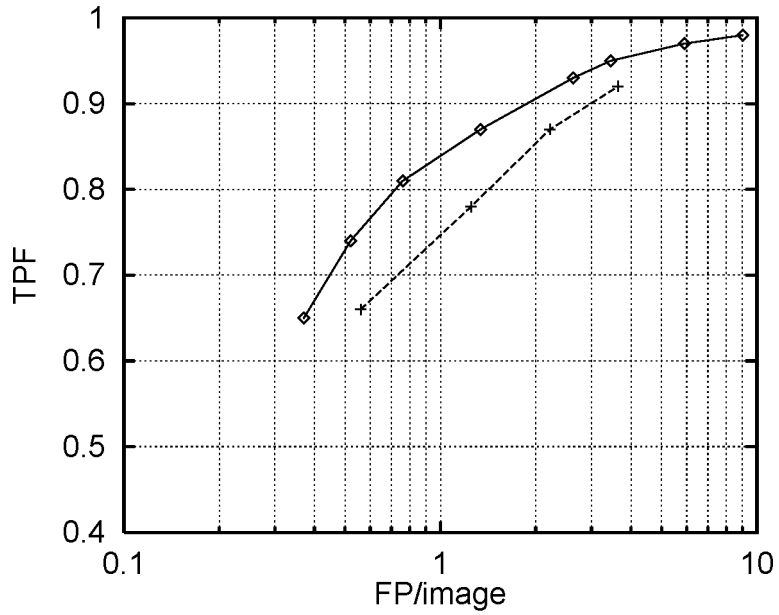


Figure 3.6: FROC curves showing the influence of using an adaptive or a fixed signal dependent noise correction, on detection of microcalcification clusters and obtained on a dataset of 245 mammograms. The adaptive ($\diamond$) and the fixed phantom based noise correction ($+$) are compared where the first is based on B-spline interpolation between bins, consisting of a variable number of pixels, wherein asymmetry is reduced.

Finally, the effect when the filter used for noise estimation is not exactly matched with the local contrast filter used for detection was investigated. In both cases asymmetry was reduced in histograms of local contrast. Using a matching 9×9 window gives superior results compared to using a 5×5 window for determination of local contrast. This can be explained by the fact that the noise which is relevant for microcalcifications in the detection scheme used, is better represented by the 9×9 window than the 5×5 window.

3.6 Discussion and conclusions

It has been shown that accurate detection of microcalcifications clusters is highly dependent on proper normalization of local contrast. In this study, a robust method for normalization of local contrast features is presented and evaluated using a large database of 245 digitized mammograms. It is shown that relatively small adjustments in the algorithm have strong influence on the detection performance. An adaptive approach is used, in which for each image at hand high frequency noise is determined as a function of the grey level. This is achieved by dividing the grey scale in bins and estimating high frequency noise from the histogram of local contrast in each bin. Noise as a function of the grey level can then be obtained and from this information local contrast features are normalized.

We investigated two variants for dividing the grey scale into bins. It appeared that detection results related to noise estimation in relatively small bins could be improved by

using much larger bins consisting of a variable number of pixels per bin M and B-spline interpolation. The value M was chosen small for the highest brightness values because these regions consist of relatively few pixels and $\sigma_c(y)$ decreases fast for increasing high pixel values. In this way we were able to obtain more reliable estimates of σ_c in these regions. At lower brightness, where $\sigma_c(y)$ is more constant, larger values for M were used. As a result noise estimates are more accurate due to bins containing a large amount of pixels.

Also, it was noticed that histograms of local contrast are asymmetrical around zero in higher and lower bins which hampers accurate noise estimation. We experimented with a number of approaches for local contrast normalization and investigated their effect on histogram asymmetry and detection performance. First we assumed μ_c to be zero and, as a consequence, neglected asymmetry. Then we corrected local contrast by taking both the standard deviation as the mean of local contrast into account. This approach appears to provide better results than the first method. Showing that asymmetry should not be neglected. Finally we investigated a method that uses a blurred image to estimate local contrast distributions. It appeared that this leads to more symmetric histograms of local contrast. This latter method gives a substantial improvement in detection results, although its performance is close to the normalization method that includes estimates of $\mu_c(k)$.

It is shown that an adaptive noise equalization gives much better results than a fixed noise equalization, probably due to the fact that noise characteristics are mammogram dependent caused by variation of film type and film development characteristics. This result confirms earlier studies performed on a much smaller public database of 40 mammograms with microcalcification [6].

3.7 Acknowledgments

The authors wish to thank the Dutch Cancer Society for supporting this work by grant KUN 95-897. Furthermore, the authors like to greatly acknowledge prof. Michael Brady from Oxford University (Medical Vision Laboratory, Engineering Science) for helpful discussion and suggestions.

Bibliography

[1] N Karssemeijer, J T Frieling, and J H Hendriks. Spatial resolution in digital mammography. *Invest radiol*, 28(5):413–9, May 1993.

[2] H Roehrig, E Krupinski, and T Yu. Physical and psychophysical evaluation of digital systems for mammography. *SPIE 2436*, pages 124–134, 1995.

[3] R M Nishikawa, M L Giger, K Doi, C J Vyborny, and R A Schmidt. Computer-aided detection of clustered microcalcifications: an improved method for grouping detected signals. *Med Phys*, 20(6):1661–1666, 1993.

[4] H P Chan, K Doi, C J Vyborny, R A Schmidt, C E Metz, K L Lam, T Ogura, Y Wu, and H Macmahon. Improvement in radiologist's detection of clustered microcalcifications on mammograms. *Inv Radiol*, 25:1102–1110, 1990.

[5] W P Kegelmeyer and M C Allmen. Dense feature maps for detection of calcifications. In A G Gale, S M Astley, D R Dance, and A Y Cairns, editors, *Digital Mammography*, pages 3–12. Elsevier, Amsterdam, 1994.

[6] N Karssemeijer. A stochastic model for automated detection of calcifications in digital mammograms. *Image and Vision Computing*, 10:369–375, 1992.

[7] T Netsch and H-O Peitgen. Scale space signatures for the detection of clustered microcalcifications in digital mammograms. *to appear in IEEE Transactions on Medical Imaging*, 18:774–786, 1999.

[8] A Maitournam, C Graffigne, and A Strauss. Modeling of digital mammograms using bicubic spline functions and additive noise. In N Karssemeijer, MAO Thijssen, JHCL Hendriks, and LJTO van Erning, editors, *Digital Mammography*, pages 177–184. Kluwer, Dordrecht, 1998.

[9] S Brown, R Li, L Brandt, L Wilson, G Kossoff, and M Kossoff. Development of a multi-feature cad system for mammography. In N Karssemeijer, MAO Thijssen, JHCL Hendriks, and LJTO van Erning, editors, *Digital Mammography*, pages 189–196. Kluwer, Dordrecht, 1998.

[10] R N Strickland and H I Hahn. Wavelet transforms for detecting microcalcifications in mammograms. *IEEE Transactions on Medical Imaging*, 15(2):218–228, 1996.

[11] W Qian, M Kallergi, L P Clarke, H D Li, P Venugopal, D Song, and R A Clark. Tree structured wavelet transform segmentation of microcalcifications in digital mammography. *Med phys*, 22(8):1247–54, Aug 1995.

[12] W Qian, L P Clarke, M Kallergi, and R A Clark. Tree-structured nonlinear filters in digital mammography. *IEEE Trans. Med. Imag.*, 13:25–36, 1994.

[13] W Richardson. Nonlinear filtering and multiscale texture discrimination for mammograms. *Proc SPIE 1768, Mathematical Methods Med. Imag.*, pages 293–305, 1992.

[14] A F Laine and S Song. Multiscale wavelet representation for mammographic feature analysis. *Proc SPIE 1768, Mathematical Methods Med. Imag.*, pages 306–316, 1992.

[15] P M Taylor. *Computer assisted decision making for image understanding in medicine.* PhD thesis, University of London (University college), 1998.

[16] W Veldkamp and N Karssemeijer. Improved correction for signal dependent noise applied to automatic detection of microcalcifications. In N Karssemeijer, MAO Thijssen, JHCL Hendriks, and LJTO van Erning, editors, *Digital Mammography*, pages 169–176. Kluwer, Dordrecht, 1998.

[17] D H Ballard and C M Brown. *Computer vision.* Prentice-Hall, Englewood Cliffs, NJ, 1982.

[18] A K Jain. *Fundamentals of digital image processing.* Prentice-Hall, Englewood Cliffs, NJ, 1989.

Chapter 4

An Improved Method for Detection of Microcalcification Clusters in Digital Mammograms[1]

Abstract

In this study it is shown that the performance of a statistical method for detection of microcalcification clusters in digital mammograms, can be improved substantially by using a second step of classification. During this second step, detected clusters are automatically classified into true positive and false positive detected clusters. For classification the k-nearest neighbor method was used in a leave-one-patient-out procedure. The sensitivity level of the method was adjusted both in the first detection step as in the second classification step. The Mahalanobis distance was used as criterion in the sequential forward selection procedure for selection of features. This primary feature selection method was combined with a classification performance criterion for the final feature selection. By applying the initial detection at various levels of sensitivity, various sets of false and true positive detected clusters were created. At each of these sets the classification can be performed. Results show that the overall best FROC performance after secondary classification is obtained by varying sensitivity levels in both the first and second step. Furthermore, it was shown that performing a new feature selection for each different set of false and true positives gave different feature sets. A large database of 245 digitized mammograms with 341 clusters was used for evaluation of the method.

[1]This chapter is based on the publication: W.J.H. Veldkamp, N. Karssemeijer, *An improved method for detection of microcalcification clusters in digital mammograms*, SPIE Medical Imaging 1999.

4.1 Introduction

Previously, a statistical method for detection of microcalcifications in digital mammograms was developed at our institute [1]. It became evident that the detection performance depends strongly on a preprocessing step of noise equalization. In the preprocessing step, local contrast features used for detection of microcalcifications, were adaptively normalized to remove signal dependency of noise. In recent work [2], normalization of local contrast was further optimized which improved detection performance substantially. In this study the detection scheme was extended by a second step of classification. During this step, detected clusters were automatically classified into true positive and false positive detected clusters. While the previous method was entirely based on classification of pixels using local features, in the secondary step various specific cluster and microcalcification properties could be modeled. In this way an improvement in the overall detection performance was expected. The overall scheme is illustrated by Fig. 4.1.

Microcalcification detection schemes based on an initial detection step followed by a second step of classification were developed by a number of researchers [3, 4, 5, 6, 7, 8]. In these studies classification was always applied on a single data set of false and true positive detections. This dataset was constructed by selecting a starting point at high sensitivity at the FROC curve corresponding to the initial detection. In this study we investigated the effect of applying secondary classification at a number of starting points at the initial FROC curve. It was questioned whether or not it is beneficial to choose a lower sensitivity starting point for the overall scheme's performance at lower sensitivity. This could be beneficial due to differences of statistical properties of data bases corresponding to different starting points. For instance because at high sensitivity of the initial detection, detected clusters will be overall larger and thereby contain more candidate microcalcifications. Furthermore, it was investigated whether performing a new feature selection at each different starting point would result in different feature sets.

The method used for initial detection is based on Bayesian techniques. A Markov random field model was used to model spatial relations between the labels in an iterative segmentation process. Three features were used for detection: the local contrast at two different spatial resolutions and the output of a line/edge detector. Local contrast features were normalized after having determined the noise as a continuous function of the grey value. This was done for each mammogram separately without using any information but the image data itself. Such a robust approach avoids dependency on, for instance, film type and film development characteristics.

A number of features were calculated for classification between true positive and false positive detections. For calculating reliable shape and contrast features an accurate segmentation method is essential [9]. A number of segmentation methods were investigated using images of a phantom consisting of small dots with mammographic background su-

perimposed. A method based on a Markov random field model appeared to give the best segmentation results [10]. In this work this segmentation method was used for calculating shape and contrast features of detected microcalcifications. Further, contrast was defined in a way that it is approximately independent of breast thickness and exposure level. For classification the k-nearest neighbor method was used in a leave-one-patient-out procedure.

A database was constructed by selecting all cases with microcalcification clusters from a four years period of breast cancer screening in Nijmegen and environs. The dataset used consists of 245 mammograms with 341 clusters. The mammograms were digitized at 0.05 *mm* per pixel and averaged down to 0.1 *mm* per pixel. The detection scheme was evaluated using this large database and by carrying out a FROC study.

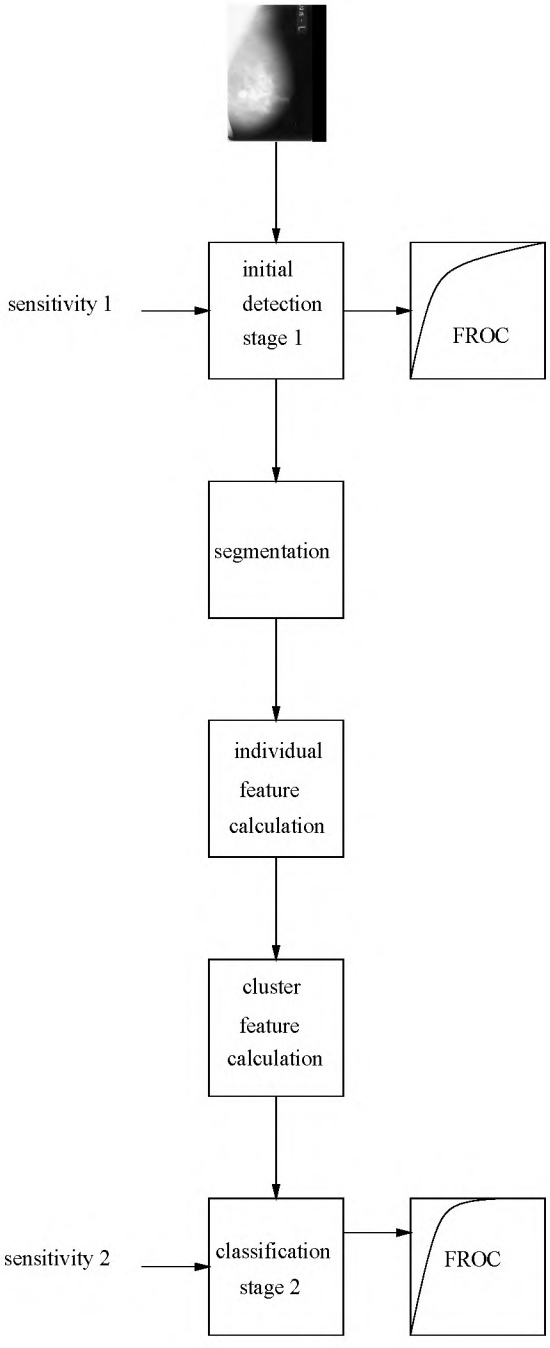


Figure 4.1: A two-stage approach for detection of microcalcification clusters.

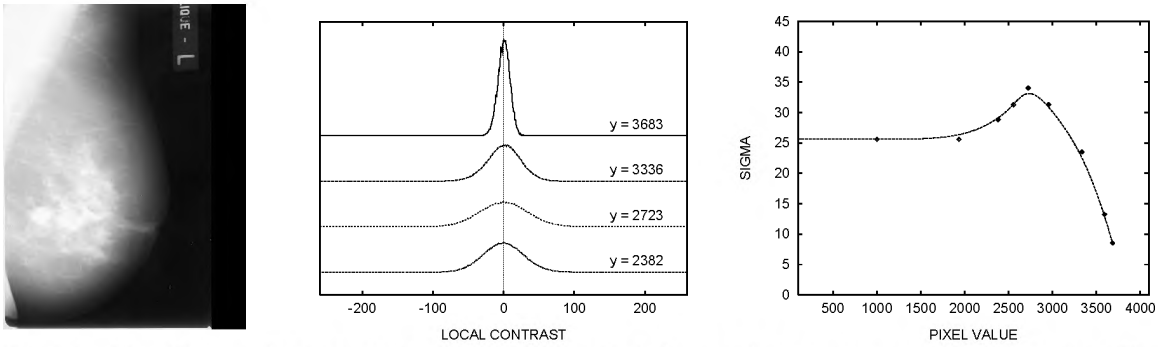


Figure 4.2: Noise characteristics were estimated for each mammogram separately. The middle figure shows histograms of local contrast for a mammogram that are determined in four different bins. Also the central pixel values of the bins are shown in the figure. At right the continuous functions $\sigma_c(y)$, obtained by interpolation of the estimates $\sigma_c(k)$, is shown.

4.2 Detection method

Detection of microcalcifications concerns deciding whether a signal is due to a calcification or to high frequency noise. In the detection scheme used, the high frequency noise relevant for detection is the uncertainty in local contrast features used for detection. High frequency noise was therefore modeled by the standard deviation of local contrast. Detection is hampered by the fact that the high frequency noise appears to vary with the grey level (Fig. 4.2; right figure). Consequently, an important step in the detection scheme is normalization of local contrast features for signal dependent noise.

4.2.1 Normalization of local contrast

Local contrast c_i at site i can be defined by:

$$c_i = y_i - \frac{1}{N}\sum_{j\in\partial_i} y_j, \tag{4.1}$$

with y_i the pixel value at site i and ∂_i a neighborhood or window at i of size N. The standard deviation of local contrast σ_c is determined as a function of grey level from the image at hand and from this information the normalization is performed. Such an adaptive approach does not rely on the stability of the image formation process and it also takes tissue inhomogeneity into account as an additional noise component.

For obtaining $\sigma_c(y)$ the grey scale is divided in non-overlapping but adjacent bins numbered $k = 1, 2, ..., K$. After computing local contrast, the probability density function $f(c|k)$ can be estimated by normalizing the histograms of c determined within each bin k. Fig. 4.2 shows an example calculated from a mammogram for four different bins. For each bin k the standard deviation $\sigma_c(k)$ of local contrast c can be calculated from $f(c|k)$. Fig. 4.2 also shows the continuous function $\sigma_c(y)$ obtained by interpolation of the noise estimates $\sigma_c(k)$.

After determination of $\sigma_c(y)$ local contrast c_i is normalized by:

$$c_i' = c_i/\sigma_c(y_i), \tag{4.2}$$

where c_i' represents normalized local contrast at site i.

We used bins of variable size where the term "size" refers to the number of pixels in a bin. Bin size is large at lower brightness where $\sigma_c(y)$ is relatively constant (Fig. 4.2; right figure). For the highest brightness values, where $\sigma_c(y)$ decreases rapidly, bin size is chosen smaller because these regions consist of relatively few pixels.

It was noticed that histograms of local contrast were asymmetrical around zero in higher and lower bins which hampers accurate noise estimation. We used a blurred image to estimate local contrast distribution. It appeared that this approach leads to more symmetric histograms of local contrast and to more accurate detection results as described extensively in [2].

4.2.2 Statistical model

The statistical model is described in detail in [1]. It is based on the use of Bayesian techniques and applications of a Markov random field model, where the latter models the fact that microcalcifications occur in clusters. The detection scheme uses three different features for representing the image data: the local contrast at two different spatial resolutions and the output of a line/edge detector. At high resolution the local contrast is determined by equation 4.1. At a lower resolution the contrast is simply calculated by smoothing the result of equation 4.2 with a 3×3 uniform filter kernel. The line/edge feature is calculated from a local histogram of gradient orientations.

During the detection process pixel labels x_i are iteratively updated by maximizing their probability, given the image data in a small neighborhood $y_{\delta i}$ of site i and given the current estimate of the rest of the labeling $\hat{X}_{S\setminus i}$:

$$x_i' = \max_l [p(x_i = l | y_{\delta i}, \hat{X}_{S\setminus i})], \tag{4.3}$$

where $l = 1,2,3,4$ represents four pixel classes: background, microcalcifications, lines/edge and film emulsion errors. The image data $y_{\delta i}$ is represented by the three local image features mentioned in the preceding subsection. The probability to be maximized can be written as

$$p(x_i = l | y_{\delta i}, \hat{X}_{S\setminus i}) \propto f(\Theta_i | x_i = l, \hat{X}_{S\setminus i})\, p(x_i = l | \hat{X}_{S\setminus i}), \tag{4.4}$$

where Θ_i is a vector denoting the values of the three features at a particular site. The *a priori* probability $p(x_i | \hat{X}_{S\setminus i})$ of the labels represents the Markov random field and models spatial relations [1]. For instance, a pixel is more likely to be part of a calcification if there are other calcifications in the neighborhood. In section 4.4 it is described how we used a Markov random field for segmentation of microcalcifications.

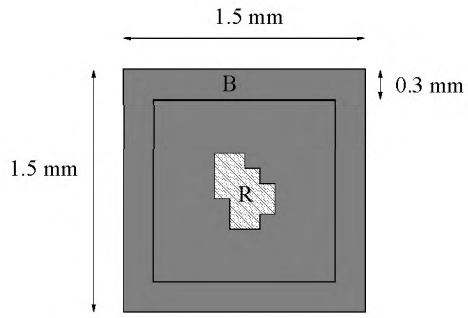


Figure 4.3: Each detected microcalcification and its local background is stored in 15 × 15 pixel data fields.

4.3 Representation of data

In this section we will explain the data structure used in this study. Detected microcalcifications and part of their neighborhood are stored in 15 × 15 pixel data fields. The calcification's center of mass, according to the annotation produced by the detection program, coincides with the data fields center (Fig. 4.3). The outer rim B of 3 pixels in width, is used for obtaining background information in the segmentation procedure and in calculating features. If the calcification area R overlaps with B, the corresponding pixels are excluded in obtaining background information. It should be noted that microcalcifications are typically smaller than 1 *mm* and will therefore rarely overlap with region B.

A number of shape and contrast features are calculated from these data fields after the segmentation step, described in the following section, is applied. Classification of true and false positive detected clusters is based on cluster features describing specific cluster properties. These cluster features are partly derived from the individual microcalcification features. For each detected microcalcification and cluster, data structures are constructed that contain the specific features. Calculation of features will be explained in section 4.5.

A header is linked to each pixel data field and each data structure of features. The headers contain, for instance, information concerning position of the microcalcification or cluster in the mammogram and information necessary for performing a FROC analysis.

4.4 Segmentation

Shape and contrast features of microcalcifications are often used in schemes for automated differentiation between true positive and false positive detected microcalcifications. In determining such features, segmentation plays an important role and influences classification and thereby detection performance [9]. For instance, overestimation (or underestimation) of object size deteriorates contrast measurement significantly [10]. By segmentation we mean here determination of the precise outline of a microcalcification. The segmentation program performs a segmentation on each of the data fields constructed as described in the previous

section. It is important to note that the original annotation, given by the detection program, is not involved in the segmentation procedure. It is assumed, however, that the center of the data field is part of the detected calcification.

To perform an accurate segmentation on detected microcalcifications a background trend correction is applied. As stated before, we assume that a detection step has been carried out so that the positions of the microcalcifications are known. To define a background area a disc with a diameter of 1 *mm* is used. Microcalcifications will be covered by this disc. The center of the disc coincides with the center of the data field. The pixel values in the disc area R' are replaced by new pixel values interpolated from the surrounding background. Pixel value y_i, with $i \in R'$ is replaced by y_i' according to the following weight function

$$y_i' = \frac{\sum_{k \in B} y_k / d_{ik}}{\sum_{k \in B} \frac{1}{d_{ik}}}, \tag{4.5}$$

where d_{ik} is the Euclidean distance between site i and site k, with k a pixel on the boundary L of R'. The background image is low-pass filtered using a 5×5 uniform kernel and then subtracted from the original image. In the second stage, the segmentation is performed on the resulting image. This image is also used for calculating contrast features of the object.

Instead of using a fixed disc to cover the object to be segmented, an initial segmentation could be used for creating a background image. However it turned out that low contrast objects can not be segmented in a consistent way, since background level and signal strength estimates are influenced by background structures. This may cause segmentation of high intensity background structures or sincere underestimation of the objects. Therefore, a background trend correction based on an initial segmentation does not give convenient results for low contrast microcalcifications. Calcifications that are, according to the original annotation, larger than the disc do not undergo the segmentation procedure. They keep their original segmentation as produced by the detection algorithm.

In an earlier study a number of segmentation methods was investigated. A distinction was made between thresholding and an iterative approach. It was found that an iterative method based on a Markov random field model gave the most accurate detection results. In a Markov Random Field (MRF) for each pixel a neighborhood is defined. The MRF model is specified by giving the conditional probability distribution of a pixel label l, with $l = 0, 1$ (background and foreground), given its grey level and the labels of its neighbors. A Gaussian model is used for representing the fluctuation of grey levels due to noise. The method is described extensively in [10].

4.5 Features

This section describes the various features that are used in the classification scheme. The classification of true and false positive detections is based on cluster features describing

specific cluster properties. These cluster features are derived from features of the individual microcalcifications composing the cluster and from their locations. Subsection 4.5.1 will describe the features corresponding to individual microcalcifications which we will indicate by the term local features. Subsection 4.5.2 describes the cluster (or global) features.

4.5.1 Individual microcalcification features

The following local features were considered for each individual detected microcalcification:

1. perimeter, defined as the number of pixel sides that touch a background pixel (a pixel is represented by a square).
2. area, represented by the number of microcalcification pixels.
3. compactness, defined as $c = \frac{perimeter^2}{4\pi area}$.
4. eccentricity, defined as $e = \frac{I_{xx}+I_{yy}-\sqrt{(I_{xx}-I_{yy})^2+4I_{xy}{}^2}}{I_{xx}+I_{yy}+\sqrt{(I_{xx}-I_{yy})^2+4I_{xy}{}^2}}$ where I_{xx}, I_{xy} and I_{yy} are the moments of inertia.
5. thickness, calculated as the width of the best fitting rectangle [11].
6. orientation, defined as the angle of axis of the least moment of inertia [11] with respect to the xy-plane.
7. direction, calculated as the relative direction in which the microcalcification is located viewed from its cluster's gravity center.
8. line, the mean of the output of the line/edge detector in a detected microcalcification.
9. background, the mean intensity level of the background.
10. foreground, the mean intensity of the detected microcalcification.
11. distance, the distance to the closest neighbor calcification.
12. The contrast measure C_i^{mc} for a microcalcification pixel at site i, depends on microcalcification thickness d_{mc}, an estimate of the linear attenuation coefficients of microcalcifications μ_{mc} and background tissue μ_b, the film curve gradient c_1 and the digitization constant c_0. In earlier work ([10]) the following expression for microcalcification pixel contrast is derived given that log-scaling is used in the digitization process:

$$C_i^{mc} = y_i - y_b = c_0 c_1 (\log e) d_{mc}(\mu_{mc} - \mu_b) \quad for\ y_i \in R. \tag{4.6}$$

From equation 4.6 it is clear that using proper scaling, microcalcification pixel contrast is approximately independent of breast thickness, and exposure level. For each pixel in a microcalcification, contrast is calculated. Several features related to the distribution of pixel contrast in a microcalcification are determined:

- maximum, mean, average deviation, skewness and kurtosis.

4.5.2 Cluster features

Cluster features are mainly related to the distribution of the local features (subsection 4.5.1) in the detected clusters. When we consider $\mathbf{l} = (\mathbf{l_1}, ..., \mathbf{l_k})^{\mathbf{T}}$ as a vector containing the k local features for describing individual microcalcifications, we can describe the following cluster features as follows:

1. The mean value of local feature l_i in a cluster, indicated as m(l_i).
2. The standard deviation of l_i in a cluster, indicated as s(l_i).
3. The minimum value of l_i in a cluster, indicated as min(l_i).
4. The maximum value of l_i in a cluster, indicated as max(l_i).

For the orientation and direction features we only calculated the standard deviations. These angles depend on the orientation of the breast in the xy-plane. The standard deviation of the angles, however, gives information that is invariant for breast positioning. We are planning to use the nipple as a reference point in future work in order to overcome dependency on orientation of the breast in the mammogram. Apart from the cluster features mentioned above, the following two supplementary cluster features are defined:

1. cluster area, represented by the number of pixels in a cluster.
2. number, defined as the number of calcifications in a cluster.

4.6 Feature selection

We used a sequential forward selection procedure which is a fast method for obtaining a list of valuable feature combinations. This procedure is designed to search for feature combinations that minimize some error function. This is equivalent to searching for features that maximize a distance measure between distributions. In this study we used the Mahalanobis distance measure as the criterion to be maximized. The method starts with selection of the feature corresponding to the largest Mahalanobis distance. In each of the following cycles an additional feature is selected from the remaining features in the pool, as the one that gives the largest contribution to the Mahalanobis distance. In this way, a list of features is

obtained from which we can determine the best combinations of k features for classification by taking the first k features in the list.

It should be noted that the Mahalanobis distance monotonically increases with an increasing number of features. However, increasing the number of features, some of which may be redundant or irrelevant, may deteriorate classification performance. The best feature combination, out of those generated by the sequential forward selection, was determined by examining the corresponding FROC curves. A criterion was used that measures the area under a predefined part of the curve. Feature combinations corresponding with a maximum area value were searched for. We actually calculated the area under the curve starting in the interval of FP/image that ranges from 0 to the level the primary detection is tuned at, since secondary classification will generate a curve between these two points.

4.7 Classification

In Fig. 4.1 the classification scheme is shown. For obtaining a FROC curve we used the leave-one-patient-out method. In this approach, all clusters from all views related to the same patient as the cluster to be classified are left out from the training set. The training set always contained the same predefined number of true and false positive detected clusters. The level of sensitivity in the first stage detection was adjusted by varying the *a priori* probability of a pixel being part of a microcalcification in the Markov random field model. In stage 2 sensitivity was adjusted by considering the ratio of false positives and true positives in the k neighbors of a input cluster. If the number of true positives out of k neighbors exceeded a threshold a cluster was classified as true. By varying the threshold a FROC curve was obtained.

The knn-method happens to be sensitive for differences in value ranges of features causing unpredictable classification behavior. This is due to the fact that the knn-method uses a distance measure in feature space for classification. Defining $\mathbf{g}$ as a vector containing the cluster features, the influence of heterogeneity in feature value ranges was limited by rescaling the cluster features g_i according to:

$$g_i' = \frac{g_i - \mu_i}{\sigma_i}, \tag{4.7}$$

with g_i' the rescaled features.

Since we can adjust the sensitivity both in the first as in the second stage classification, it is not obvious how to combine the two sensitivity levels in order to obtain an optimal FROC curve. In our experiments we first obtained a FROC curve for the first detection step only. Subsequently we performed the second stage classification on sets of false and true positive clusters corresponding to different sensitivity levels in the first stage detection. By varying the sensitivity at the second stage as well, a number of new FROC curves were obtained

from the initial detection FROC curve. In principle, an overall curve could be computed by selecting the maximum true positive fraction for each number of false positives per image.

4.8 Performance evaluation

For performance evaluation the numbers of true and false positive clusters were determined for each mammogram, while the sensitivity was varied. A FROC (free response operating characteristics) curve can be constructed by plotting the true positive fraction as a function of the number of false positive clusters per image. Labeling of the true clusters was done according to the radiologic screening report.

For counting the number of true positives, a cluster was regarded as detected if two or more calcifications were found in the marked area covering the cluster. No verification of the individual calcifications was performed. With respect to the false positives, a cluster was counted if a closed area was found in which two or more calcifications occurred. The closed area has to be enclosed by an empty region of 0.5 *cm* in width. Clusters of large benign calcified cysts, benign fat necrosis calcifications and benign vascular calcifications are not regarded as true clusters in this study.

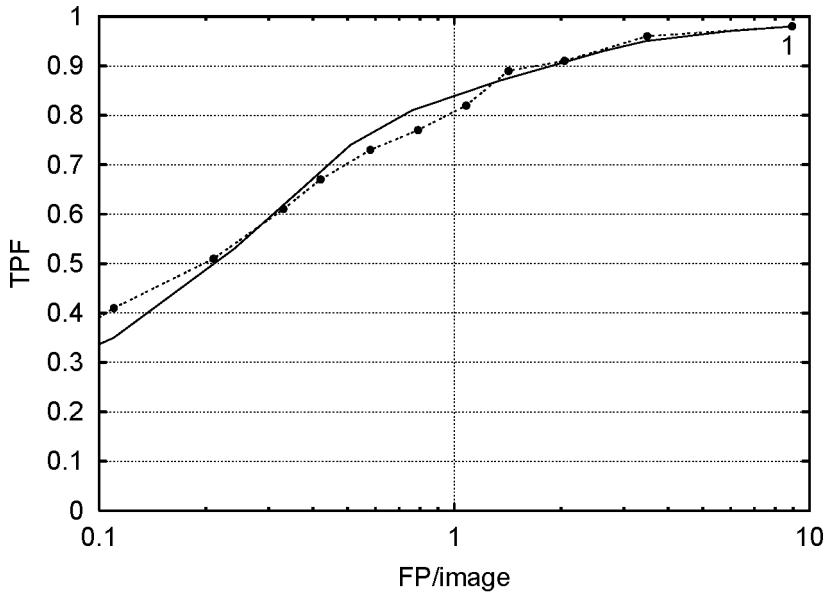


Figure 4.4: Results of first stage detection (solid line) and second stage classification (•). Second stage classification is performed at sensitivity level 1 at the first stage detection curve.

4.9 Results

In Fig. 4.4 and 4.5 second stage classification results are shown. We investigated three different starting points at the original detection curve. Each point corresponds with a data

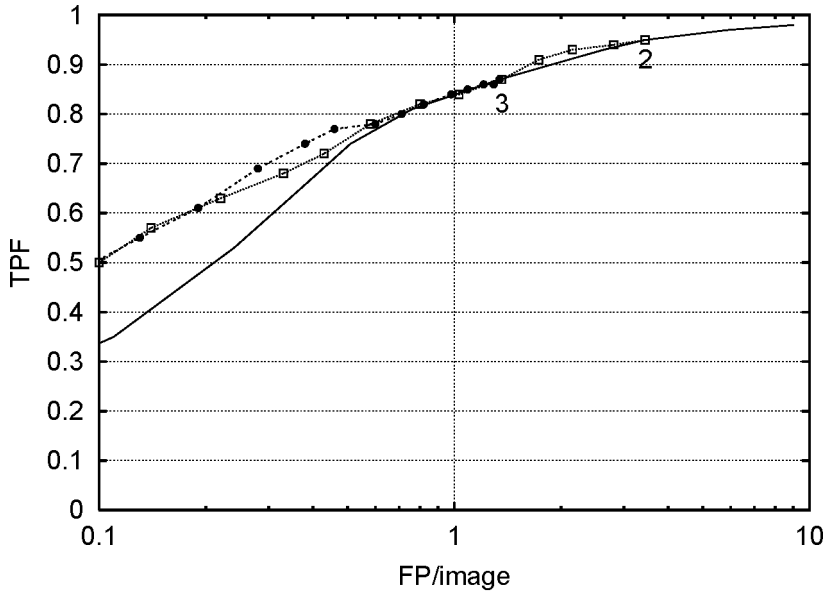


Figure 4.5: First stage detection (solid line) and second stage classification results. Second stage classification is performed at two different sensitivity levels at the first detection curve. The related curves are starting at the points 2 (□) and 3 (•).

Rank	Level 1	Level 2	Level 3
1	cluster area	cluster area	min(line)
2	max(foreground)	m(distance)	max(foreground)
3	m(distance)	min(line)	s(direction)
4	min(background)	max(foreground)	m(kurtosis)
5	min(line)	min(background)	s(mean)
6	m(skewness)	s(kurtosis)	max(eccentricity)
7	m(thickness)	m(line)	m(line)
8	m(line)	s(direction)	s(eccentricity)
9	max(eccentricity)	s(maximum)	m(eccentricity)
10	s(eccentricity)	max(eccentricity)	max(background)
11	m(eccentricity)	s(eccentricity)	s(thickness)
12	m(foreground)	m(eccentricity)	max(line)
13	m(background)	max(background)	s(orientation)
14	m(mean)	m(compactness)	m(distance)
15	max(line)	m(average deviation)	m(area)

Table I: The first 15 features that were selected by the sequential forward selection.

base of true and false positive detected clusters and for each of these data bases a feature selection is performed. At each starting point 15 selected features were used as shown in Table I. Using the area criterion, we found for all starting points that the results did not improve when using more than 15 features. In the range of 15 to 20 features results were

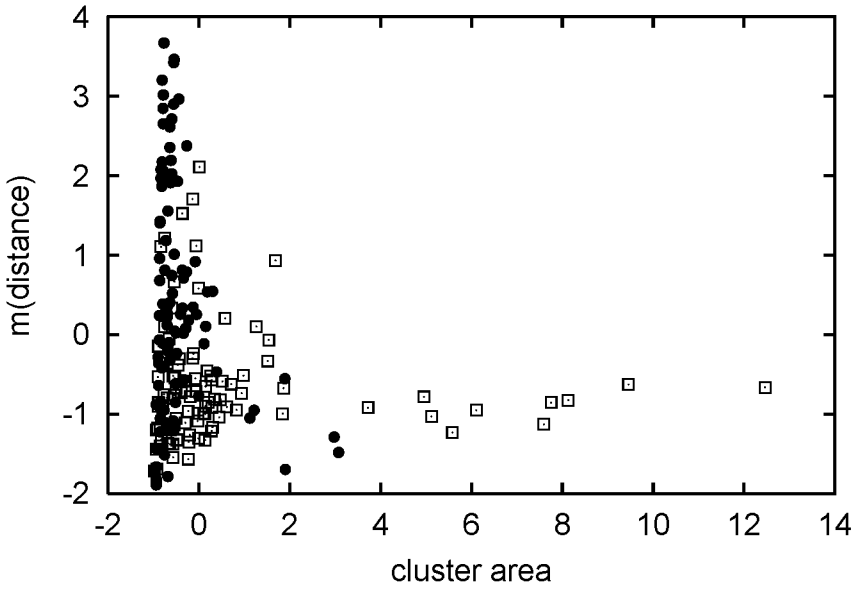


Figure 4.6: Scatter plot corresponding to cluster area and min(distance) (i.e. the minimal distance in a cluster between two calcifications) at level 2. The features are scaled by subtracting the mean and division by the standard deviation. The true positive clusters (□) and false positive clusters (•) are shown.

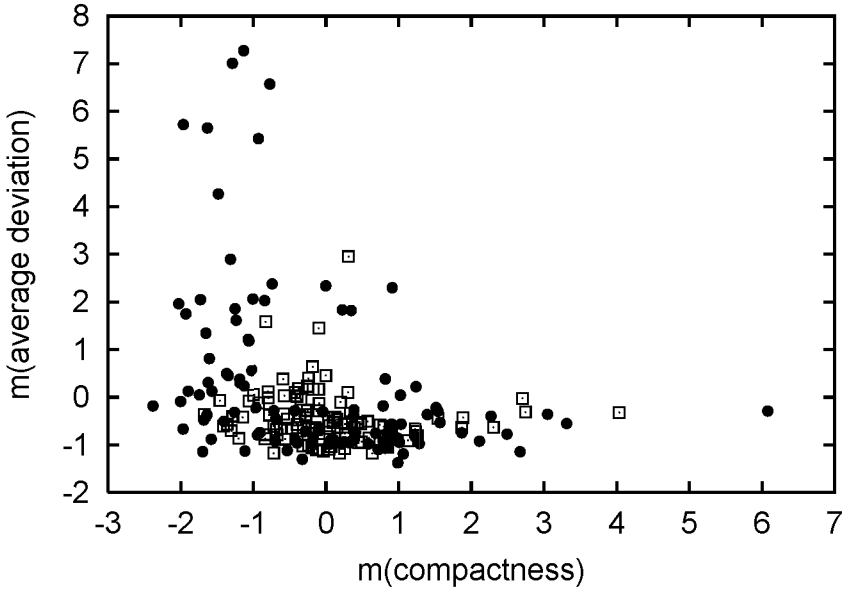


Figure 4.7: Scatter plot corresponding to m(compactness) and m(average deviation) at level 2. The features are scaled by subtracting the mean and division by the standard deviation. The true positive clusters (□) and false positive clusters (•) are shown.

stable. Outside this range results deteriorated. Clearly, in Fig. 4.4 the curve related to secondary classification starting at high sensitivity (point 1) does not show an improvement compared to the original detection curve. It follows from Fig. 4.5 that the curves starting at level 2 and 3 respectively, outperform the original curve for sensitivity values below 0.8.

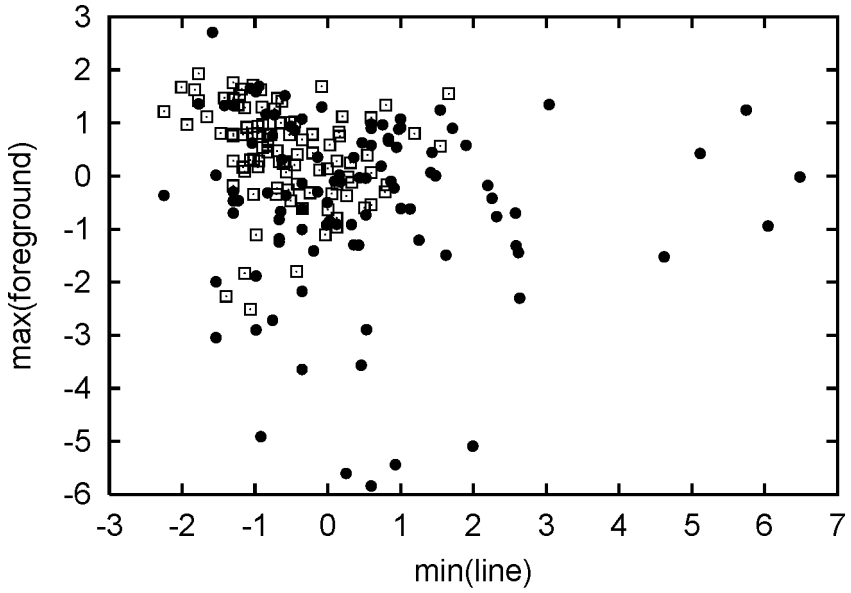


Figure 4.8: Scatter plot corresponding to min(line) (i.e. the minimum line value) and max(foreground) (i.e. the maximum foreground pixel value) at level 3. The features are scaled by subtracting the mean and division by the standard deviation. The true positive clusters (□) and false positive clusters (•) are shown.

For instance, at a true positive fraction of 0.5, the number of false positive detected clusters is roughly halved. Furthermore, it appears that starting secondary classification at level 2 or 3 gives comparable results. Table I gives the best feature combinations found in relation to the three starting points. It follows that the feature sets corresponding to sensitivity levels 1 and 2 show important overlap. It appears that the min(line) feature becomes more important for decreasing sensitivity levels. On the other hand, the cluster area is not selected at level 3 whereas it appears to be the most important feature at levels 1 and 2.

In Figs. 4.6, 4.7 and 4.8 scatter plots are shown that give insight in the discriminative power of features at two different sensitivity levels. An equal number of points is selected for both classes in the plots. Furthermore, the features are scaled according to equation 4.7. In Fig. 4.6 the two most valuable features, according to Table I, for sensitivity level 2 are plotted. True clusters appear to have larger cluster area and have smaller m(distance) values at this sensitivity level. Fig. 4.7 gives an impression of the discriminative power of features selected in cycle 14 and 15 with respect to level 2. The features m(compactness) and m(average deviation) are plotted here. The variation in pixel contrast appears to be larger in a number of false positive detected clusters. Fig. 4.8 shows the first two selected features at level 3. It becomes clear that the max(foreground) value of false positive clusters is often lower due to the fact that these detections are more often located outside the dense area of the breast. Furthermore, a substantial part of the false positive clusters shows larger values for the min(line) feature.

4.10 Discussion and conclusions

In this study it was shown that the performance of a statistical method for detection of microcalcification clusters in digital mammograms can be improved substantially by using a second step of classification. A number of features were calculated for classification of true positive and false positive detections. For calculating reliable shape and contrast features of microcalcifications, an accurate segmentation method was used based on a Markov random field model. A sequential forward selection procedure was applied for obtaining valuable feature combinations. The most accurate feature combinations, out of those generated by the sequential forward selection, were determined by examining the corresponding FROC curves. A large data set was used for evaluation of the overall detection scheme. This data set consists of 245 mammograms with 341 microcalcification clusters.

In this paper different data sets of false and true positive detected clusters were constructed by selecting different starting points at the FROC curve belonging to the initial detection. We investigated the effect of applying secondary classification at a number of these starting points. It was found that it is beneficial to perform classification starting from a point at lower sensitivity for the overall scheme's performance at lower sensitivity. Furthermore, it appeared that although feature sets selected at the various starting points show overlap, they show differences in composition as well. This can be explained by the fact that the data bases of true and false positive detections corresponding to the various starting points have specific statistical properties. For instance, because at high sensitivity of the initial detection, clusters will be overall larger and thereby contain more detections. True positive clusters at high sensitivity may even contain a lot of false positive detected microcalcifications. This limits the discriminative power of features that are based on individual microcalcification features.

From feature selection it followed that at high sensitivity the cluster area is an important feature, due to a substantial number of true positive detected clusters of relative large size. For all sensitivity levels investigated, it appeared that false positive clusters are more associated with line/edge structures. Results suggest that for classification, line/edge related features become more important at lower sensitivity starting points. Furthermore, the distance between microcalcifications appears to be smaller in true positive clusters. Finally, false positive clusters are more often located outside the dense area of the breast resulting in lower foreground and background intensity values. For future work, we are planning to use a segmentation algorithm that gives a better probability measure of pixels being located in fibro glandular tissue.

ACKNOWLEDGMENTS

The authors wish to thank the Dutch Cancer Society for supporting this work by grant KUN 95-897.

Bibliography

[1] N Karssemeijer. Adaptive noise equalization and detection of microcalcification clusters in mammography. *Int. Journal of Pattern Recognition and AI*, 7(6):1357–1375, 1993.

[2] W Veldkamp and N Karssemeijer. Improved correction for signal dependent noise applied to automatic detection of microcalcifications. In N Karssemeijer, MAO Thijssen, JHCL Hendriks, and LJTO van Erning, editors, *Digital Mammography*, pages 169–176. Kluwer, Dordrecht, 1998.

[3] R M Nishikawa, Y Jiang, and M L Giger et al. Performance of automated cad schemes for the detection and classification of clustered microcalcifications. In A G Gale, S M Astley, D R Dance, and A Y Cairn, editors, *Digital Mammography*, pages 13–20. Elsevier, Amsterdam, 1994.

[4] C S Carman and G Eliot. Detecting calcifications and calcification clusters in digitized mammograms. In K Doi, M L Giger, R M Nishikawa, and R A Schmidt, editors, *Digital Mammography*, pages 253–255. Elsevier, Amsterdam, 1996.

[5] T Hara, K Hirako, H Fujita, T Endo, K Horita, M Ikeda, C Kido, and T Ishigaki. Automated detection algorithm for clustered microcalcifications based on density gradient and triple-ring filter analysis. In K Doi, M L Giger, R M Nishikawa, and R A Schmidt, editors, *Digital Mammography*, pages 257–262. Elsevier, Amsterdam, 1996.

[6] P Sajda, C D Spence, J C Pearson, and R M Nishikawa. Integrating multi-resolution and contextual information for improved microcalcification detection in cad. In K Doi, M L Giger, R M Nishikawa, and R A Schmidt, editors, *Digital Mammography*, pages 291–296. Elsevier, Amsterdam, 1996.

[7] D Meersman, P Scheunders, and D van Dyck. Classification of microcalcifications using texture-based features. In N Karssemeijer, M A O Thijssen, J H C L Hendriks, and L J T O van Erning, editors, *Digital Mammography*, pages 233–236. Kluwer, Dordrecht, 1998.

[8] R F Nagel, R M Nishikawa, J Papaioannou, and K Doi. Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms. *Medical physics*, 25(8):1502–1506, 1998.

[9] W J H Veldkamp and N Karssemeijer. Influence of segmentation on classification of microcalcifications in digital mammography. In *18th Annual international conference of the IEEE engineering in medicine and biology society*, 1996.

[10] W J H Veldkamp and N Karssemeijer. Accurate segmentation and contrast measurement of microcalcifications in mammograms: A phantom study. *Med Phys*, 25:1102–1110, 1998.

[11] A K Jain. *Fundamentals of digital image processing*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

Chapter 5

Automated Classification of Clustered Microcalcifications into Malignant and Benign Types[1]

Abstract

The objectives in this study were to design and test a fully automated method for classification of microcalcification clusters into malignant and benign types, and to compare the method's performance with that of radiologists. Novel aspects of the approach are that the relative location and orientation of clusters inside the breast was taken into account for feature calculation. Furthermore, correspondence of location of clusters in *MLO* and *CC* views, was used in feature calculation and in final classification.

Initially, microcalcifications were automatically detected by using a statistical method based on Bayesian techniques and a Markov random field model. For classification a method based on two classification steps was developed. In the first step, classification of clusters was performed and in the second step a patient based classification was done. A total of sixteen features was used in the study. To identify meaningful features, a feature selection was applied, using the area under the ROC curve (A_z value) as a criterion. For classification the *k*-nearest-neighbor method was used in a leave-one-patient-out procedure. A database of 192 mammograms with 280 true positive detected microcalcification clusters was used for evaluation of the method. The set consisted of cases that were selected for diagnostic work up during a four years period of screening in Nijmegen and environment. Because of the high positive predictive value in the screening program (50%), this set did not contain obvious benign cases. The method's best patient-based performance on this set corresponded with $A_z = 0.83$, using nine features.

A subset of the data set, containing mammograms from 90 patients, was used for comparing the computer results to radiologists' performance. Ten radiologists read these cases

[1]This chapter is based on: W.J.H. Veldkamp, N. Karssemeijer, J.D.M. Otten, J.H.C.L. Hendriks *Automated classification of clustered microcalcifications into malignant and benign types*, Submitted to Medical Physics.

on a light-box and assessed the probability of malignancy for each patient. All participants had experience in clinical mammography and just underwent a two-weeks training session for becoming certified for screening. Results on the subset showed that the method's performance ($A_z = 0.83$) was considerably higher than that of the radiologists ($A_z = 0.63$).

5.1 Introduction

Clustered microcalcifications are an important mammographic sign of early (in situ) breast cancer. However, several benign diseases show microcalcifications as well. Mammography is the most sensitive method for early detection of breast cancer but its specificity for differentiating malignant and benign microcalcification clusters is relatively low. However, in nation-wide screening programs, accurate characterization of microcalcification clusters is essential, because recalling all cases with microcalcification clusters would result in too many false positives. This is because around 80% of all clusters appearing in the screening population are due to benign processes. Recalling all women that have microcalcifications would cause anxiety and thereby discourage women to participate in the program. Also costs of screening would increase considerably. In general, computer-aided diagnosis (CAD) is considered to be one of the most promising approaches that may improve the efficacy of mammography [1].

Only few works are related to computer-aided differentiation of benign and malignant lesions. Most researchers used either human extracted features ([2],[3]), or manual identification of calcifications along with computer extraction of features as the input for their automatic classification system. For instance, Jiang et al. described a method based on the latter approach ([4]). In their method eight features were used, based on important signs used by radiologists. On a dataset of 100 mammograms from 53 patients the best patient-based performance of their method corresponded to an area under the ROC curve of $A_z = 0.92$ where the mean performance of five radiologists gave $A_z = 0.89$. The method was applied to clusters and microcalcifications that were annotated and located by expert radiologists. Another method for classification of microcalcification clusters was developed by Chan et al. [5]. They initially used a number of morphological features. By adding various texture measures they were able to improve the classification performance substantially. Microcalcifications were annotated by expert radiologists. The best performance of their method corresponded with an A_z value of 0.89 using a data set of 145 mammographic microcalcification clusters. The fact that texture measures gave substantial information is attributed by the authors to texture changes in the breast tissue due to a developing malignancy.

Applying automated analysis (or classification) at microcalcifications that are annotated or located by radiologists avoids problems related to detection of false positive or false negative microcalcifications. However, such approach hampers clinical use of automated in-

terpretation of microcalcification clusters. Schmidt et al. addressed this draw-back and they developed a fully automated identification and interpretation method for microcalcification clusters [6]. They used a data set of 100 patients containing microcalcification clusters with known histology. They found that their computer system could not infer a reliable diagnosis with respect to cases rated by human experts as being hard to diagnose (atypical cases). Therefore, automatically detected microcalcification clusters were classified by an artificial neural network into typical and atypical clusters. Finally, only the typical clusters were classified by a third network into malignant and benign types.

In this work a fully automated method was developed for classification of microcalcification clusters into malignant and benign types. Microcalcifications were detected automatically in cluster regions that were annotated according to radiologic screening reports. Specific features with respect to the detected candidate microcalcifications within the marked area were calculated and used for classification. In order to detect microcalcifications in faint or vague clusters, the sensitivity level of the detection method had to be high which introduced false positive detected microcalcifications as well in our classification method.

For calculating reliable shape and contrast features an accurate segmentation method appeared to be essential [7]. A method based on a Markov random field was used for segmentation of the detected microcalcifications [8]. Also the effect on the classification performance when using three other more straightforward segmentation methods was investigated.

For classification of benign and malignant clusters, sixteen features were calculated to represent each cluster. We considered distribution features (based on the distribution of microcalcification properties within a cluster), cluster shape features and cluster position features. To define the latter two types of features, we used the location of the nipple and the pectoral muscle in the mammograms. A fully automated method was used for finding the pectoral muscle [9], and for determining the approximate location of the nipple.

In this study, we experimented with a classification method consisting of two classification steps. The first-step classifier, using the *k*nn-method, was used for classification of clusters in both *CC* and *MLO* views. In the second classification step the outputs of the first-step-classifier were used to obtain a patient based classification result. It was investigated whether a heuristic method that intends to link corresponding clusters in both views, would improve results. The output of the linking method was used both in patient-based classification as in feature calculation.

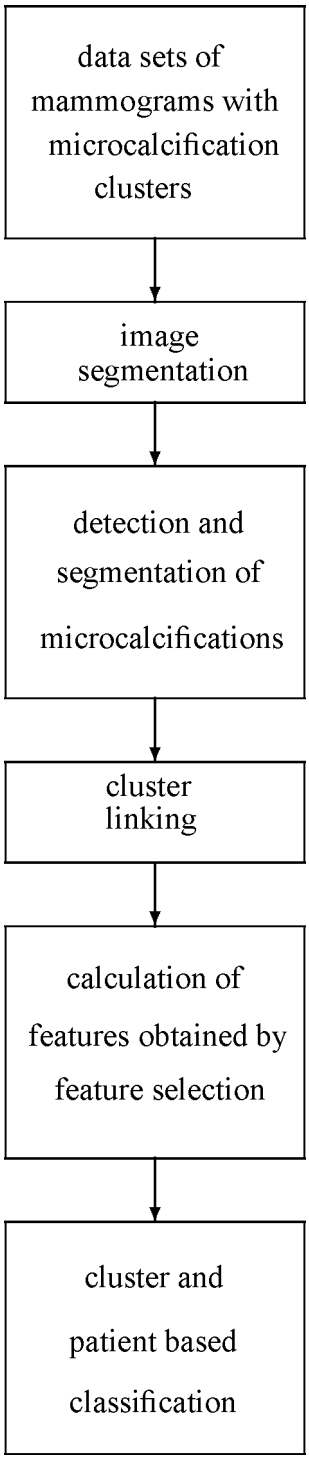


Figure 5.1: A flow chart representation of the method.

5.2 Materials and Methods

5.2.1 Image dataset

A data set was constructed by collecting all cases with reported microcalcifications that were selected by radiologists for follow-up examination during four years of screening in Nijmegen and environment. This initial data set contained four cases that were histologically classified as lobular carcinoma in situ (LCIS). When LCIS is found no further treatment is performed because it is generally regarded as a benign disease. On the other hand, one believes that LCIS forms a risk factor for developing non-comedo DCIS [10]. In our opinion, LCIS forms a third type to be classified (besides malignant and benign types) but the number of four cases was too small to train the classifier. Therefore, we decided to leave the four cases related to LCIS out of the data set. Furthermore, we wanted to focus on microcalcifications only. Therefore we excluded all cases that had mass signs as well. This resulted in a final dataset consisting of 192 mammograms from 104 women with 280 detected clusters. In all but 8 women the pathology was verified by open surgical biopsies. Women for whom no biopsy was taken have been followed for at least four years to exclude the possibility that a malignancy was present. Since the positive predictive value (*PPV*) with respect to referral of cases with microcalcification clusters in the Dutch screening program is around 0.5, obvious benign clusters were not represented in this set. The mammograms were digitized at 0.05 *mm* pixel-size (12 bits per pixel) using a Lumisys 85 digitizer and averaged down to 0.1 *mm* per pixel.

5.2.2 Image segmentation

An overview of the complete method is given in Fig. 5.1. At first an image segmentation was applied which subdivides mammograms into three distinct area's: breast tissue, pectoral muscle and background. The pectoral muscle is normally visible in the *MLO* views. To segment the background from tissue area, a global thresholding technique was applied. The threshold was determined automatically from the histogram of pixel values computed over the whole mammogram. For detection of the pectoral muscle, a region of interest was determined automatically containing the pectoral muscle. A technique based on application of the Hough transform was used to locate the position of the pectoral in this region of interest [9].

For this work, determination of nipple location was added to the segmentation procedure described above. Since the nipple is often hardly visible in a mammogram, we used knowledge concerning the geometry of the breast to determine the point at the border of the breast area that should be close to the nipple. In the *MLO* view we determined a point at the skin-line which has largest distance to the pectoral muscle. Determination was done within a range of $3\,cm$ in vertical direction, centered around the center of mass of the breast tissue

Figure 5.2: Segmentation of a mammogram. The breast area, pectoral muscle and nipple location are depicted in the segmentation result image (right).

region. In the *CC* view the same procedure was followed but now a point at the skin-line with largest distance to the chest side of the mammogram was determined. The approach appeared to be accurate enough for the purpose of this study. If necessary, in future work local features could be used to improve the accuracy. Fig. 5.2 gives an impression of the segmentation that resulted in a mask image giving breast area, pectoral muscle, nipple location and background area. The estimated nipple location is depicted by the small white square.

5.2.3 Microcalcification detection and segmentation

The microcalcification detection method was based on Bayesian techniques and application of a Markov random field model, where the latter modeled the fact that microcalcifications occur in clusters [11]. The detection scheme used three different features for representing the image data: the local contrast at two different spatial resolutions and the output of a line/edge detector. For this study a sensitivity level was used where microcalcification clusters in all mammograms of the 104 women were found. In only one film no cluster was detected. The detection scheme was previously evaluated using a database of 245 image and by carrying out FROC studies [12, 13].

Detected candidate microcalcifications and part of their neighborhood were stored in 1.5×1.5 *mm* pixel data fields. The calcifications center of mass, according to the segmentation produced by the detection program, coincided with the data fields center. An outer rim, of 3 pixels in width, was used in each data field for estimating a background level in the segmentation procedure and in calculating features.

To calculate features, a clustering procedure was performed within the annotated cluster region. Clustering was done by defining discs of 1 *cm* in diameter, centered at each detected spot. By taking groups of discs that touch or overlap cluster area's were defined. In case more than one closed area was found in the annotated region, the largest area was selected and used for calculating features. In almost all cases our procedure to define a cluster produced one cluster per annotation.

5.2.4 Cluster linking

The distance to the pectoral muscle, was one of the features in our classification scheme. This feature can only be determined for clusters in the oblique view since the pectoral muscle is in general not visible in the cranio caudal view. However, by linking clusters in the *MLO* view with corresponding clusters in the *CC* view we assigned the distance feature, being calculated for clusters in the *MLO* view, to corresponding clusters in the *CC* view as well.

Correspondence between clusters was also used in classification of patients. For this purpose, the classification results of corresponding clusters were combined to obtain a patient based result.

Linking of lesions in different views was investigated by Good et al. [14]. They used five features for linking suspicious regions in two views (*CC* and *MLO*) and found the distance between a cluster and the nipple as the most effective measure to determine clusters correspondence.

We assume that the best way to link clusters in *CC* and *MLO* views is the one in which the sum of the difference in the distances for a set of pairs s_i in *CC* and *MLO* is as small as possible, taking all possible sets s_i of pairs into account (with $s_i \in S$ where S represents all possible sets of pairs). This approach is expressed by equation 5.1. Here $D_{cc}(p)$ is the distance to the nipple for cluster p in the *CC* view, whereas $D_{mlo}(p, s_i)$ is the distance to the nipple of cluster p in the *MLO* view give s_i.

$$s_{min} = \arg\min_{s_i \in S} \{ \sum_{p} (D_{cc}(p) - D_{mlo}(p, s_i))^2 \}, \tag{5.1}$$

A problem is that an equal number of clusters does not always exist in the two views due to projection. For instance, clusters that are near can be projected as one cluster in one view where they may appear as two or more in the other view. For this reason, implementation of equation 5.1 becomes rather complex. The problem is that the set S of possible pairs is not well defined. Therefore, we used a modified procedure described in the next paragraph, which uses the same idea of minimizing the differences in distance to the nipple.

For a given pair of mammograms a $m \times n$ matrix ΔD was constructed by calculating the squared differences in distance ΔD_{pq} for all pairs of possibly corresponding clusters p and q and with m and n the numbers of clusters in both views respectively. The expression

for ΔD_{pq} is given by equation 5.2.

$$\Delta D_{pq} = (D_{cc}(p) - D_{mlo}(q))^2, \tag{5.2}$$

After the matrix ΔD had been constructed, the linking procedure started with searching the minimum difference value ΔD_{kl} in the matrix. Then, clusters k and l were linked and excluded from the linking procedure. The procedure was repeated until all clusters in at least one view were linked with different clusters in the other view. When the number of clusters m in one view was larger than the number of clusters n in the other view, a number of $m-n$ clusters in this first view, remained initially unlinked. To obtain correspondence for each cluster, each of these $m-n$ clusters was linked to that (already linked) cluster in the other view that gave the smallest difference in distance to the nipple.

5.2.5 Features

In this study, a total of sixteen features were used for classification of microcalcification clusters. We distinguished three feature types: distribution features (based on the distribution of individual microcalcification features within a cluster), cluster shape features (for instance cluster area and cluster eccentricity) and cluster location features (describing the location of clusters in a mammogram). With the definition of these features we tried to represent descriptions that are used by radiologists as described in section 5.2.5.

Features used by radiologists

With respect to benign microcalcifications, one can distinguish lobular microcalcifications (calcified milk and sclerosing adenosis) and involution microcalcifications. For malignant processes with microcalcifications, ductal carcinoma in situ (DCIS) is the most important. One defines well differentiated, intermediately differentiated, and poorly differentiated DCIS [15], where the latter type is most aggressive.

Among the important characteristics reported by radiologists are:

- Polymorphism versus monomorphism: malignant microcalcifications tend to be polymorph whereas benign cluster are more often characterized by monomorphous calcifications of uniform size [16].

- Size: some benign types of calcifications are larger and have more contrast than malignant calcifications.

- Branching type versus round and oval type: linear calcifications may be an indication of DCIS. This is due to the fact that calcifications associated with DCIS are located in the glandular ducts. Round and oval calcifications are often located in the lobuli and are often due to benign diseases.

- Orientation: malignant clusters and their calcifications, that are associated with DCIS often have shapes that are oriented to the nipple [16].
- Number: in case a cluster consists of only few microcalcifications, the cluster is regarded as less suspicious.
- Location: lesions located in the outer upper quadrant are suspicious because 48% of the cancerous processes are located in this quadrant [17].

Features used in the method

Feature type	Feature	Symbol
Distribution features	number of microcalcifications in a cluster	N
	mean microcalcification contrast C	$m(C)$
	mean microcalcification area a	$m(a)$
	mean microcalcification eccentricity e	$m(e)$
	mean microcalcification compactness c	$m(c)$
	mean microcalcification orientation α	$m(\alpha)$
	st. dev. of microcalcification contrast C	$s(C)$
	st. dev. of microcalcification area a	$s(a)$
	st. dev. of microcalcification eccentricity e	$s(e)$
	st. dev. of microcalcification compactness c	$s(c)$
	st. dev. of microcalcification orientation α	$s(\alpha)$
Cluster shape features	cluster area	A
	cluster orientation	θ
	cluster eccentricity	E
Cluster position features	relative distance to pectoral edge	d_{pm}
	relative distance to breast edge	d_{be}

Table I: The sixteen descriptions used in this study were based on typical marks as reported by radiologists.

Table I gives an overview of the features used in this study. These features were based on radiologists' description, as given in section 5.2.5. We distinguished distribution features, cluster shape features and cluster location features. The cluster center, that was used in calculating some of these features, was calculated as the center of mass, regarding all microcalcification pixels within the cluster.

Distribution features

We denote $\mathbf{l} = (l_1, ..., l_k)^T$ as a vector containing k local features for describing individual microcalcifications. Cluster features that we used, representing the distribution of local features within a cluster, were the mean m(l_i) and the standard deviation s(l_i).

Features representing individual microcalcifications are described below. Three microcalcification shape features that we investigated were area, compactness and eccentricity [7]. Microcalcification area a was simply calculated by counting the number of pixels in a segmented calcification. Compactness of a microcalcification was defined as $c = \frac{p^2}{4\pi a}$, with p the calcification's perimeter and a it's area. The eccentricity was calculated as

$$e = \frac{I_{xx} + I_{yy} - \sqrt{(I_{xx} - I_{yy})^2 + 4I_{xy}^2}}{I_{xx} + I_{yy} + \sqrt{(I_{xx} - I_{yy})^2 + 4I_{xy}^2}}, \tag{5.3}$$

where I_{xx}, I_{xy} and I_{yy} are the moments of inertia.

For calculating microcalcification contrast C, a relation was derived in [8]. In approximation, the contrast measure C is proportional to $d_{mc}(\mu_{mc} - \mu_b)$, where d_{mc} is microcalcification thickness and μ_{mc} and μ_b are the linear attenuation coefficients of microcalcifications and background tissue respectively. It should be noted that C is approximately independent of breast thickness and exposure level. To normalize the contrast we divided C by microcalcification thickness which is calculated as the width of the microcalcification's best fitting rectangle [18].

For calculating microcalcification orientation α we used a line filter as described in [19]. At a given level of spatial scale σ (in this study we took $\sigma = 0.2mm$), convolution of each mammogram in the data set was performed using three filter kernels $W_\sigma(\theta_n)$ that are second order directional derivatives of a Gaussian kernel $G(r;\sigma)$ where $G(r;\sigma) = \frac{1}{2\pi\sigma^2}\exp(-\frac{r^2}{2\sigma^2})$ and with $\theta = n\pi/3$ and n=0,1,2. This was sufficient to make an accurate operator for determining line orientation. For definition of orientation features we computed for each pixel in a segmented microcalcification the difference between θ and the orientation of a line pointing from the microcalcification at hand to the nipple. As a result these angles were always in the interval $[0, 1/2\pi]$. The relative microcalcification orientation α was determined by calculating the mean difference angle over all microcalcification pixels.

Cluster shape features

The cluster area A was defined as the area of the closed cluster region relative to the area of the breast region (where the pectoral muscle was excluded).

Cluster eccentricity E was calculated analogous to microcalcification eccentricity e in equation 5.3. For calculating cluster eccentricity, only microcalcification pixel sites were taken into account within the cluster region. In addition, cluster orientation θ was defined as the angle of axis of the clusters least moment of inertia [18] (again, considering only microcalcification pixel sites within the cluster region). The orientation was taken relative to the nipple by taking the angle between a virtual line through the cluster center with orientation θ and a line through the nipple and the cluster center.

Cluster position features

According to [17], lesions located in the outer upper quadrant are more suspicious because 48% of the cancerous processes are located in this quadrant. Furthermore, in [20] mammograms of 86 women under 50 years of age with mammographically detected cancers were reviewed. It was found that 73% of the cancers were located at the periphery of the breast. This periphery is defined as a zone 1 *cm* wide beneath the subcuteneous fat or anterior to the retromammary fat. For this reason, clusters located close to the pectoral muscle might be more likely to be due to cancerous processes. The relative distance between a cluster and the pectoral muscle d_{pm} was calculated as the shortest distance from the cluster center to the pectoral muscle divided by the distance between the nipple and the pectoral muscle. This feature could only be determined in the *MLO* view. Using cluster linking, we could also assign this distance feature to clusters in *CC* views.

Additionally, we used the distance d_{be} of a cluster to the breast edge. Here, d_{be} was taken relative to the distance from the chest-wall to the nipple.

5.2.6 Feature selection

For feature selection we used a sequential forward selection procedure. Feature selection was basically used in this study to get insight in the number of features needed to obtain useful classification results. In general a feature selection is performed by maximizing or minimizing some criterion that indicates the discriminative power of features. Our selection method used the area under the ROC-curve A_z as a criterion to be maximized.

We started with a pool consisting of the sixteen cluster features described in section 5.2.5. First, the feature that gave the largest discriminative power according to the selection criterion was searched for. This feature was then removed from the pool and added to a list. In each of the following cycles, the feature was selected that gave the largest contribution to the features already selected. The order in which they appeared in the list gave an indication of usefulness.

5.2.7 Classification

The classifier we used is shown in Fig. 5.3. The classifier is trained using clusters originating from both *MLO* and *CC* views. Each cluster is represented by a vector $\mathbf{g}$ containing its features. For each patient, these vectors are presented to the first classifier. As a result of this first classification step, each cluster is assigned a likelihood l^c. The final patient based classification result l^p was computed in two ways. Firstly, it was taken equal to the maximum cluster likelihood l^c_{max} during the second classification step. In this case the most suspicious cluster determines the likelihood of malignancy of the patient. In the second approach, the classifier combined classification results of corresponding clusters in *MLO*

and *CC* views by taking their mean likelihood of malignancy $\bar{l}^c$. The final patient based classification result l^p was taken equal to the maximum mean cluster likelihood $\bar{l}^c_{max}$ found in the patient.

In the first step classifier, the *k*-nearest-neighbor (*k*nn-) method was used which is a relatively simple but fast classification method. For each vector to be classified the same number of randomly selected benign and malignant training vectors was used. All other vectors related to the same patient were excluded from the training set. A cluster likelihood of malignancy l^c was taken equal to the ratio of malignant vectors among the *k* nearest neighbors. Overall classification results were obtained by using the leave-one-patient-out method. By varying a threshold in the classifier with respect to the cluster or patient likelihood of malignancy (l^c or l^p, respectively) ROC curves could be constructed. We used Metz's software ([21]) for obtaining ROC curves. For the computer performance the LABROC4 program was used.

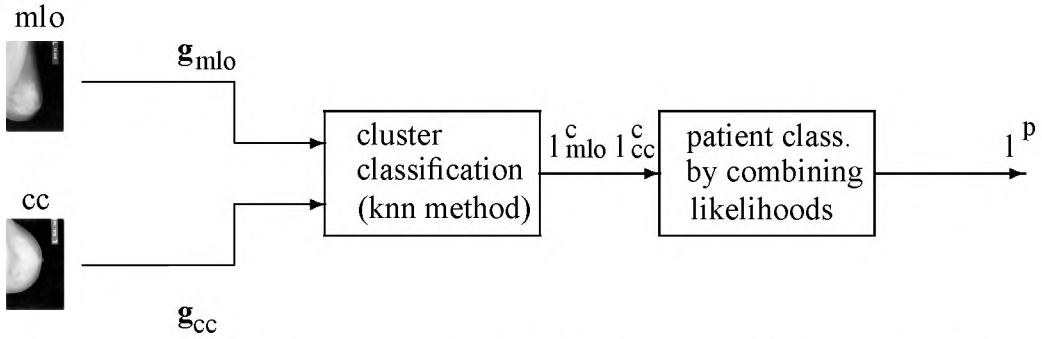


Figure 5.3: Cluster and patient based classification. An example is illustrated where two clusters (one in each view) are involved.

5.2.8 Observer study

In order to compare the computer performance with human performance an observer study was carried out. Ten radiologists participated in this study. Each observer had at least a 4 years period of experience in clinical mammography. Observers were recruited when they were on a two-weeks educational visit at the National Expert and Training Center for Breast Cancer Screening at the University Hospital of Nijmegen. By taking this course, radiologists become certified to participate as screenings radiologists in the Dutch screening program. The radiologists were asked to participate during the last days of their visit.

A subset of the data set, containing mammograms from 90 patients, was used for comparing the method's performance to human performance. The films from the remaining patients were not available in the archives at the time of the study. The original films of these 90 cases were presented on a light box. The order of cases was randomly chosen but mammograms of one patient were always presented together. For each case to be judged, a

Rank	Feature	Symbol	A_z value clusters	A_z value patients
1	relative distance to pectoral edge	d_{pm}	0.56	0.57
2	relative distance to breast edge	d_{be}	0.65	0.71
3	st. dev. of microcalcification area a	$s(a)$	0.67	0.71
4	st. dev. of microcalcification orientation α	$s(\alpha)$	0.70	0.75
5	st. dev. of microcalcification contrast C	$s(C)$	0.69	0.77
6	mean microcalcification area a	$m(a)$	0.72	0.80
7	mean microcalcification orientation α	$m(\alpha)$	0.72	0.82
8	cluster area	A	0.70	0.80
9	number of calcifications	N	0.73	0.83
10	cluster orientation	θ	0.69	0.77
11	st. dev. of microcalcification compactness c	$s(c)$	0.68	0.77
12	mean of microcalcification eccentricity e	$m(e)$	0.67	0.75
13	mean of microcalcification compactness c	$m(c)$	0.69	0.78
14	cluster eccentricity	E	0.67	0.74
15	mean of microcalcification contrast C	$m(C)$	0.65	0.74
16	st. dev. of microcalcification eccentricity e	$s(e)$	0.62	0.71

Table II: The order of features in which they were selected by the sequential forward selection using to the A_z criterion.

scorings sheet was available. The images were printed on the sheet together with the annotations of the clusters. The radiologists assessed a probability of malignancy for each patient based on what they found, the most suspicious cluster in the patient. The radiologists were asked to judge the cases on a confidence rating scale ranging from certainly benign to certainly malignant. ROC curves demonstrating classification performance were generated for each observer using ROCFIT [21].

5.3 Results

At first we determined an appropriate value for the number of neighbors k, in the $k$nn-method. We compared classification performances of using sixteen features for different values of k. It was found that taking k=15 gave best results where experiments were carried out for k with values 40, 20, 15, 10 and 5. For k=5 and k=40 the performance dropped substantially. For the other values of k results were similar. The features were rescaled in order to limit the influence of heterogeneity in feature value ranges as described in [22].

Table II shows results of feature selection. It gives A_z values for cluster based and patient based classification. From this table it appears that the distance of a cluster to the pectoral muscle was the most valuable feature. In this section, results of patient classification were obtained by classification using the maximum mean likelihood of malignancy

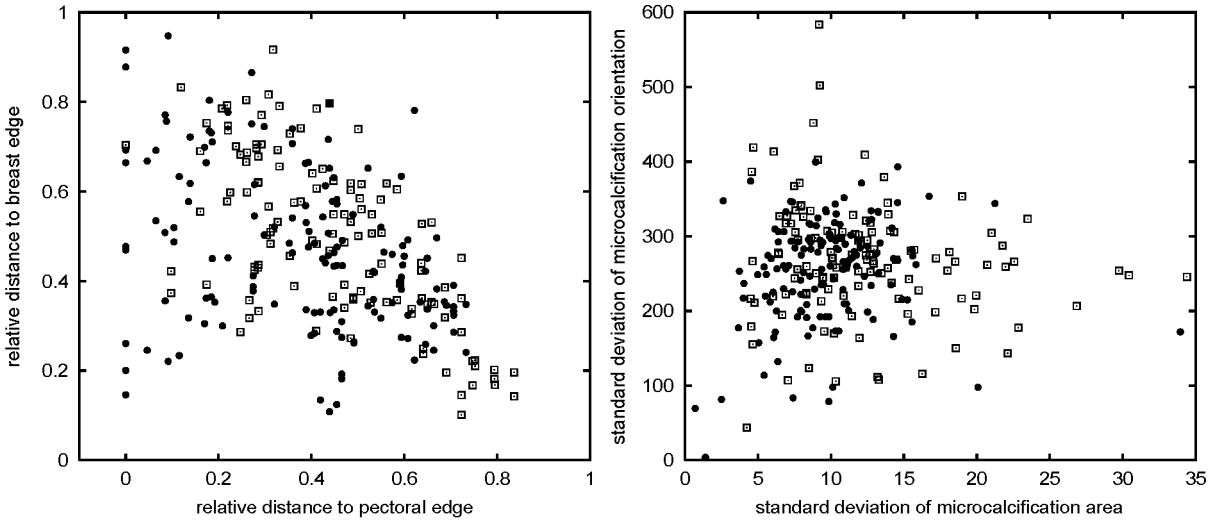


Figure 5.4: The left scatter plot shows the relative distance to the pectoral edge versus the relative distance to the breast edge. The right plot depicts standard deviations of microcalcification area and microcalcification orientation. Benign (□) and malignant (•) clusters are shown.

of corresponding clusters (as described in section 5.2.7). It was found that this approach gave superior results compared to taking the maximum cluster likelihood in classification of patients. For instance, the former approach gave A_z = 0.83 (patient classification) using the nine first selected features, whereas the latter approach gave A_z = 0.76 using the same features.

Fig. 5.4 gives scatter plots of (non-rescaled) features. The left scatter plot shows the relative distance to the pectoral edge versus the relative distance to the breast edge. The distance to the pectoral of clusters that were located in the pectoral region was taken equal to zero. It appears that clusters close to the pectoral or close to the breast edge (values close or equal to zero) are more likely to be malignant (•). The number of clusters in the data set near the nipple (values close to one) appear to be benign (□). The right plot depicts standard deviations of microcalcification area and microcalcification orientation. Benign clusters show larger standard deviation of microcalcification area.

Fig. 5.5 shows ROC-curves obtained by using the first nine features from Table II. The right plot shows results for patient based classification (A_z=0.83; curve *A*), whereas the left plot shows results for cluster-based classification (A_z=0.73; curve *A*). In both plots the upper bound (curve *C*) and lower bound (curve *B*) of the 95% confidence interval are shown.

In Fig. 5.6 results on a subset of 90 patients are shown. Computer results using three (A_z=0.69; *B*), six (A_z=0.81; *C*) and nine (A_z=0.83; *D*) features are shown. Also the mean performance of ten radiologists is depicted in the figure (A_z=0.63; *A*). For obtaining this latter curve, the scoring data of the radiologists was pooled. The computer performances were obtained by using the total data set in a leave-one-patient-out procedure, identically as in obtaining the previous results.

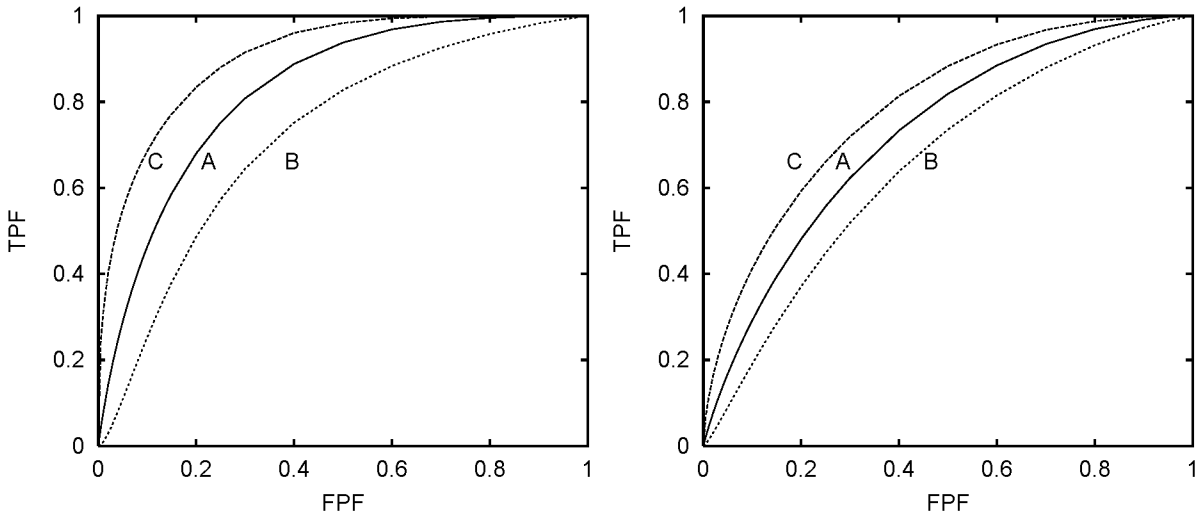


Figure 5.5: Computer classification performance using nine features. ROC curves for cluster based classification (left; curve *A*) and patient based classification (right; curve *A*) are shown. In both plots, the upper bound (curve *C*) and lower bound (curve *B*) of the 95% confidence interval are shown.

Apart from the iterative microcalcification segmentation method (based on a Markov random field) used in the experiments above, we investigated three other segmentation methods. These three methods were based on thresholding: one method used a noise dependent threshold, one method used a fixed contrast level and one method used a signal dependent threshold [8]. When using the first nine features in Table II, the iterative method gave the highest A_z value. As curves related to the other methods were within the 95% confidence interval differences could not be regarded as significant.

5.4 Discussion and conclusions

In this study an automated method was developed for classification of microcalcification clusters into benign and malignant types. A data set that consisted of 192 mammograms originating from 104 women with 280 detected (true positive) microcalcification clusters, was used for evaluation of the method. It was found that the relative distance to the pectoral and the skin-line gave important information. Clusters in the data set that were located close to the pectoral muscle or breast edge were more likely to be malignant than clusters located deeper in the breast. Additional to this finding an interesting study is reported in [20]. In this study it was found that 73% of the mammographically visible breast cancers in women under 50 years old were located near the skin or pectoral muscle. The authors suggest that this phenomenon is caused by the fact that a large part of the glandular tissue is located in this zone. However, they recognize that this does not fully explain their observations.

Apart from the location features it was found that the standard deviations of microcalcification area, orientation and finally contrast are three important features. However, it

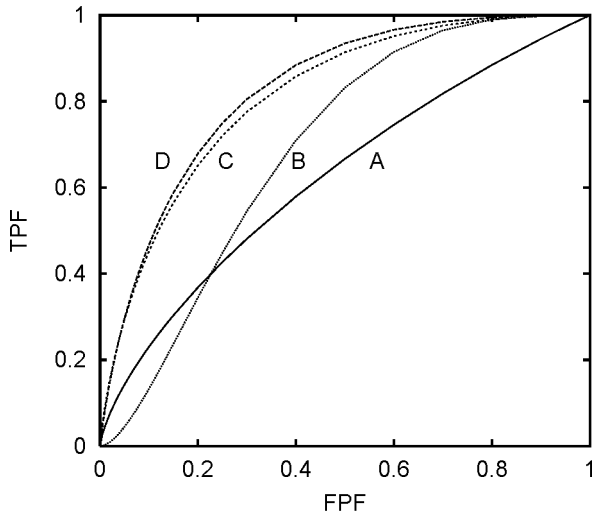


Figure 5.6: Computer classification performance using nine (*D*), six (*C*) and three (*B*) features outperforms the mean performance of ten radiologists (*A*).

was beneficial to include more features in the classification scheme where using nine features gave the best performance. The patient-based performance on this set corresponded with $A_z = 0.83$, using these nine features. The corresponding ROC curve showed that the computer could classify almost all malignant clusters correctly (sensitivity=0.995) at a false positive fraction of 0.8. This is a promising result suggesting that this CAD system could reduce the number of unnecessary biopsies by 20%.

Results on a sub set showed that the method's performance ($A_z = 0.83$) is considerably higher than that of the radiologists ($A_z = 0.63$). It was noticed that there were substantial differences in the performances of the ten radiologists. Their results were in the range of $A_z = 0.50$ to $A_z = 0.72$ (s.d. = 0.07).

Since we used only one data set for classification and feature selection we might have introduced a positive bias in the computerized classification results. In order to find out how large this bias might be, we experimented with random feature values. The feature selection procedure was kept identical to using real features but now we started from sixteen features that consisted of randomly chosen values. This experiment was repeated 20 times. It was found that patient based classification using random features, gave A_z values that were around 0.6 for using the first selected feature to using the first 10 selected features. When using 11 selected features or more, performance dropped fast to $A_z = 0.5$ (using 16 features). Using real features, performance increased steady from using 1 feature ($A_z = 0.57$) to using 9 features ($A_z = 0.83$) features. For using more features it fell to $A_z = 0.71$ (using 16 features). It can be concluded that the results we obtained when using real features may be somewhat biased, but that a large bias is unlikely. It should be noted that we selected our pool of features to represent characteristics used by radiologists, which makes bias less likely.

For future work we are planning to extend this study by including the performance of a number of expert radiologists that have longer experience in screening mammography. Furthermore, we want to evaluate the method on datasets that represent the whole spectrum of microcalcification clusters in the screening population including also false negatives of screening.

ACKNOWLEDGMENTS

The authors wish to thank the Dutch Cancer Society for supporting this work by grant KUN 95-897. Furthermore they want to thank greatly, all the radiologists who participated in the observer study.

Bibliography

[1] J Roehrig, T Doi, A Hasegawa, B Hunt, J Marshall, H Romsdahl, A Schneider, R Sharbaugh, and W Zang. Clinical results with r2 imagechecker in support of fda pma application. In N Karssemeijer, MAO Thijssen, JHCL Hendriks, and LJTO van Erning, editors, *Digital Mammography*, pages 395–400. Kluwer, Dordrecht, 1998.

[2] J A Baker, P J Korngruth, Y J Lo, M E Williford, and C E Floyd jr. Breast cancer: prediction with artificial neural network based on bi-rads standardized lexicon. *Radiology*, 196:817–820, 1995.

[3] Y Wu, M L Giger, K Doi, C J Vyborny, R A Schmidt, and C E Metz. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology*, 187:81–87, 1993.

[4] Y L Jiang, R M Nishikawa, D E Wolverton, C E Metz, M L Giger, R A Schmidt, C J Vyborny, and K Doi. Malignant and benign clustered microcalcifications: Automated feature analysis and classification. *Radiology*, 198:671–678, 1996.

[5] H P Chan, B Sahiner, K L Lam, N Petrick, M A Helvie, M M Goodsitt, and D D Adler. Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces. *Medical physics*, 25(10):2007–2019, October 1998.

[6] F Schmidt, E Sorantin, C Szepesvari, E Graif, M Becker, H Mayer, and K Hartwagner. An automatic method for the identification and interpretation of clustered microcalcifications in mammograms. *Phys. med. biol*, 44:1231–1243, 1999.

[7] W J H Veldkamp and N Karssemeijer. Influence of segmentation on classification of microcalcifications in digital mammography. In *18th Annual international conference of the IEEE engineering in medicine and biology society*, 1996.

[8] W J H Veldkamp and N Karssemeijer. Accurate segmentation and contrast measurement of microcalcifications in mammograms: A phantom study. *Med Phys*, 25:1102–1110, 1998.

[9] N Karssemeijer. Automated classification of parenchymal patterns in mammograms. *Phys Med Biol*, 43:365–378, 1998.

[10] D Dronkers, J H C L Hendriks, R Holland, and G Rosenbusch. *Radiologische mammadiagnostik*. Georg Thieme Verlag, Rüdigerstrasse 14, D-70469 Stuttgart, 1999.

[11] N Karssemeijer. A stochastic model for automated detection of calcifications in digital mammograms. *Image and Vision Computing*, 10:369–375, 1992.

[12] W Veldkamp and N Karssemeijer. Improved correction for signal dependent noise applied to automatic detection of microcalcifications. In N Karssemeijer, MAO Thijssen, JHCL Hendriks, and LJTO van Erning, editors, *Digital Mammography*, pages 169–176. Kluwer, Dordrecht, 1998.

[13] W J H Veldkamp and N Karssemeijer. Normalization of local contrast in mammograms. *to appear in IEEE Transactions on Medical Imaging*, 2000.

[14] W F Good, B Zheng, Y-H Chang, X H Wang, G Maitz, and D Gur. Multi-image cad employing features derived from ipsilateral mammographic views. *SPIE 1999 Image processing*, 3661:474–485, 1999.

[15] R Holland, J L Peterse, R R Millis, V Eusebi, D Faverly, M J van de Vijver, and B Zafrani. Ductal carcinoma in situ: a proposal for a new classification. *Seminars in diagnostic pathology*, 11(3):167–180, 1994.

[16] M Lanyi. *Breast calcifications*. Springer-Verlag, Berlin Heidelberg, 1988.

[17] J R Harris, S H Hellman, I C Henderson, and D W Kinne. *Breast diseases*. J.B. Lippincott Company, East Washington Square, Philadelphia, Pennsylvania 19105, 1991.

[18] A K Jain. *Fundamentals of digital image processing*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[19] N Karssemeijer and G M te Brake. Detection of stellate distortions in mammograms. *IEEE Trans Med Imag*, 15:611–619, 10 1996.

[20] A Stacey-Clear, K A McCarthy, D A Hall, E Pile-Spellman, G White, C Hulka an G J Whitman, E F Halpern, and D B Kopans. Mammographically detected breast cancer: location in women under 50 years old. *Radiology*, 186(3):677–680, 1993.

[21] C E Metz, J H Shen, and B A Herman. New methods for estimating a binormal roc curve from continiously-distributed test results. Presented at the 1990 Joint Statistical Meetings of the American Statistical Society and the Biometric Society, Anaheim, Calif, August 1990.

[22] W J H Veldkamp and N Karssemeijer. An improved method for detection of microcalcification clusters in digital mammograms. *SPIE medical imaging*, 3661:512–522, 1999.

Summary and conclusions

To detect breast cancer as early as possible and hence to increase the survival chance, breast cancer screening programs have been introduced in a number of countries. One of the important radiologic tasks in the Dutch screening program is detection and characterization of microcalcification clusters. Clustered microcalcification may indicate breast cancer at an early stage, but 80% of the clusters encountered in the female breast are due to benign processes.

This thesis described methods for automated detection and characterization of microcalcification clusters. It is hoped that detection and characterization methods as described in this thesis can improve the diagnostic performance of radiologists in screening mammography. For the studies described in Chapters 3,4 and 5, databases were composed using cases with microcalcification clusters from a four year period (beginning in 1991) of breast cancer screening in Nijmegen and environs. A total amount of 245 mammograms from 125 women was available. The mammograms were digitized at 0.05 *mm* per pixel linear with optical density using a Lumisys 85 digitizer and were averaged down to 0.1 *mm* per pixel.

In **Chapter 2** methods were investigated for segmentation of detected microcalcifications. Performances of different segmentation methods were compared using a phantom consisting of a pattern of dots in the range of 0.2-0.8 *mm*. These dots were taken as a model for microcalcifications. Apart from segmentation performance, also influence of the modulation transfer function on contrast estimates was determined and the effect of exposure levels on segmentation was analyzed.

An iterative method based on a Markov random field with parameters estimated locally and an approach that uses a signal dependent threshold gave satisfying results. The performances of both methods were comparable. Thresholding methods based on a local noise dependent threshold or a fixed contrast threshold gave very poor results, yielding size and contrast measurements that can hardly be used for classification purposes.

Experiments with a mammographic background added to the phantom images were carried out to investigate use of a background trend correction on segmentation. It was shown that also in mammographic images the iterative method and the signal dependent method worked well provided that a background trend correction was carried out. Additionally, it appeared that overall results were hardly affected by the exposure level. An effect of exposure level on segmentation accuracy was only observed for dots with very low contrast. In

measuring contrast a correction factor had to be applied since contrast of small objects is degraded by the MTF of the imaging chain.

Chapter 3 described correction for signal dependency of high frequency image noise. This appears to be highly important in automated detection of microcalcification clusters. A new technique was presented for dividing the grey scale in sample intervals (bins) and by using a model for describing additive high frequency noise. By estimating the noise level in each bin, noise as a function of the grey level could be adaptively determined for each image separately by only using the image information. By using noise as a function of the grey level, local contrast features used for detection were rescaled to remove signal dependency of noise.

An approach based on a variable bin size appeared to give the best detection results. Furthermore, by using a model that describes additive high frequency noise, results could be further optimized. It was shown that this optimized method for adaptive noise equalization gave substantially better detection results than a fixed noise equalization.

In **Chapter 4**, it is shown that by adding a secondary step of classification to the statistical detection method (described in Chapter 3), detection performances could be improved substantially. In the second step of classification, detected clusters were automatically classified into true positive and false positive detected clusters. By applying the initial detection at various levels of sensitivity, various sets of false and true positive detected clusters were created. At each of these sets a classification was performed. For calculating reliable shape and contrast features of candidate microcalcifications, the accurate segmentation method based on a Markov random field (Chapter 2) was used. A sequential forward selection procedure was used for feature selection. From the experiments it became clear that it is beneficial to perform classification starting from a point at lower sensitivity. This can be explained by the fact that the data bases of true and false positive detections corresponding with the various starting points have different statistical properties. For instance, primary detection at high sensitivity results in large detected clusters containing more detections. True positive clusters detected at high sensitivity may even contain a lot of false positive detections. This limits the discriminative power of features that describe individual microcalcifications. It was found that the features that describe line/edge structures as well as the distance between calcifications are important at the sensitivity levels investigated. Also pixel values of the background, indicating whether detections are located in fat or glandular tissue contribute substantially in classification performance. At high sensitivity the cluster area gives important information.

Finally, in **Chapter 5** a fully automated method is developed for characterization of microcalcification clusters. Initially, microcalcifications were automatically detected by using the statistical method that was investigated in Chapter 3. A total amount of sixteen features was used in the study. For segmentation, the iterative method based on a Markov random field was used. Also the effect on the characterization performance when using three other

more straightforward segmentation methods was investigated. Furthermore, it was investigated whether a heuristic method that intends to link corresponding clusters in two different views of the same breast, would improve results. The output of the linking method was used both in characterization as in feature calculation.

Ten radiologists read cases from a subset on a light-box and assessed the probability of malignancy for each patient. All participants had at least four years experience in clinical mammography. Results on the subset showed that the method's performance was considerably higher than that of the radiologists. Results on the larger data set even suggest that this CAD scheme could reduce the number of false positive referrals by 20%. It was found that the relative distance to the pectoral and skin line, the standard deviations of microcalcification area, orientation and contrast are important features. However, it was beneficial to include more features in the classification scheme. The heuristic method for cluster linking turned out to be useful both in calculation of location features and in improving classification results.

For future work it is planned to extend this study by including the performance of a number of expert radiologists that have large experience in screening mammography. Furthermore, we want to evaluate the method (as well as the features used in it) on datasets that represent the whole spectrum of microcalcification clusters in the screening population including also false negatives of screening.

Samenvatting en conclusies

Om de kans op sterfte aan borstkanker bij vrouwen te verkleinen, is het essentieel deze ziekte in een zo vroeg mogelijk stadium op te sporen. Om die reden zijn in verschillende landen bevolkingsonderzoeken naar borstkanker in gang gezet. In het Nederlands bevolkingsonderzoek naar borstkanker speelt het detecteren en karakteriseren van microcalcificatieclusters een belangrijke rol. Microcalcificaties zijn kleine verkalkingen die in de borst kunnen voorkomen. Grofweg variëren deze verkalkingen in diameter van 0.1 mm tot 0.5 mm. Clusters van microcalcificaties kunnen een indicatie vormen van borstkanker dat zich in een pril stadium bevindt. Echter 80% van de clusters die in de vrouwenborst worden gevonden zijn het gevolg van een goedaardige afwijking. De bedoeling van het karakteriseren van microcalcificatieclusters is een onderscheid te maken tussen maligne (kwaadaardige) en benigne (goedaardige) typen.

In dit proefschrift worden automatische methoden beschreven voor het detecteren en karakteriseren van microcalcificatieclusters. Het doel van dergelijke methoden is om de diagnostische prestatie van radiologen in het bevolkingsonderzoek te verbeteren door middel van ondersteuning met de computeroutput. De basis voor de studies in dit proefschrift werd gevormd door een dataset van 245 mammogrammen. Deze mammogrammen zijn afkomstig van 125 doorverwezen vrouwen bij wie microcalcificatieclusters zijn gevonden in het bevolkingsonderzoek naar borstkanker in Nijmegen en omgeving. Deze mammogrammen waren gedigitaliseerd met een Lumisys 85 digitizer op 0.05 mm per pixel en met pixelwaarden die lineair zijn met de optische densiteit. De beelden zijn voor de experimenten uiteindelijk teruggebracht in resolutie naar 0.01 mm per pixel.

Na de algemene inleiding van **hoofdstuk 1**, zijn in **hoofdstuk 2** methoden onderzocht voor segmentatie van gedetecteerde microcalcificaties. Segmentatie wil hier zeggen het bepalen van de precieze omtrek (begrenzing) van een object. De prestaties van een aantal segmentatiemethoden zijn met elkaar vergeleken gebruikmakend van zogenaamde fantoomopnamen. Het originele fantoom bestaat uit een patroon van stippen die in diameter variëren van 0.2 tot 0.8 mm. Deze stippen zijn als model genomen voor microcalcificaties. Naast het beoordelen van de segmentatieresultaten, werd ook de invloed van de 'modulation transfer function' (MTF) op contrastmetingen van de stippen vastgelegd, alsmede het effect van de mate van gebruikte röntgenexposie op het segmentatieresultaat.

Een iteratieve methode gebaseerd op een Markov random field, waarbij de parameters

lokaal bepaald worden ('iterative method') en een methode die gebruik maakt van een signaalafhankelijke drempel ('local signal dependent threshold'), gaven beide goede resultaten. Bovendien waren de prestaties van beide methoden vergelijkbaar. Methoden die gebruik maakten van een drempel die gekozen was afhankelijk van de lokale beeldruis ('local noise dependent threshold') of een vast gekozen drempel ('fixed contrast threshold') gaven daarentegen matige resultaten. De oppervlakte en contrastmetingen bij deze twee methoden met betrekking tot de fantoomstippen leken ongeschikt voor classificatiedoeleinden.

Eveneens werd in dit hoofdstuk het nut van een achtergrondcorrectie voor segmentatie bepaald door een mammografische achtergrond te superponeren op fantoombeelden. Ook in het geval van een achtergrondcorrectie bleken de iteratieve en de signaalafhankelijke methode naar tevredenheid te werken (wanneer getest met deze mammografische beelden.). Daarnaast bleek dat de algehele prestaties nauwelijks beïnvloed werden door de gebruikte röntgenexposie. Slechts voor de stippen met zeer laag contrast werd een effect op de segmentatie-accuratesse waargenomen. Bij het meten van contrasten moest een correctiefactor worden meegenomen omdat het contrast van kleine objecten door de MTF van het totale beeldvormende systeem negatief wordt beïnvloed.

Hoofdstuk 3 beschreef methoden om voor de signaalafhankelijkheid van hoogfrequente beeldruis te corrigeren. Dit is gebleken een zeer belangrijke stap te zijn voor het automatisch detecteren van microcalcificatieclusters. Een nieuwe techniek werd gepresenteerd die de grijswaardenschaal in sample-intervallen (bins) opdeelt en die gebruik maakt van een model dat hoogfrequente ruis beschrijft. Door in elk bin de ruis te meten, kon de ruis op een adaptieve manier als functie van de grijswaarde voor ieder afzonderlijk beeld worden bepaald, door alleen van de beeldinformatie gebruik te maken. Met behulp van ruis als een functie van de grijswaarde konden lokale contrastkenmerken ('local contrast features') die voor detectie gebruikt worden, worden genormeerd om op die manier de signaalafhankelijkheid te verwijderen.

De methode die gebruik maakte van een variabele bingrootte bleek de beste detectieresultaten te geven. Verder kwam naar voren dat het model om hoogfrequente ruis te beschrijven de resultaten nog verbeterde. Het werd duidelijk dat de geoptimaliseerde methode voor adaptieve ruisequalisatie substantieel betere detectieresultaten gaf dan een vaste ruisequalisatie.

In **hoofdstuk 4** werd getoond dat door toevoeging van een classificatieprocedure aan de statistische detectiemethode (zoals beschreven in hoofdstuk 3), de detectieresultaten aanzienlijk verbeterden. In deze classificatieprocedure werden gedetecteerde clusters automatisch geclassificeerd in 'true positive' en 'false positive' gedetecteerde clusters. Door de initiële detectie op verschillende niveaus van sensitiviteit toe te passen, werden verschillende sets van 'false' en 'true positive' gedetecteerde clusters gecreëerd. Op elk van deze sets kon een classificatie worden uitgevoerd. Om betrouwbare vorm- en contrastkenmerken te berekenen van kandidaat-microcalcificaties, werd de iteratieve segmentatiemethode ge-

bruikt die gebaseerd is op een Markov random field (Hoofdstuk 2). Een sequentieel voorwaartse selectiemethode is gebruikt voor selectie van kenmerken. Het bleek gunstig uit te werken om de classificatie toe te passen op sets die corresponderen met een lagere sensitiviteit. Dit kan verklaard worden door het feit dat datasets die bij verschillende niveaus van sensitiviteit horen, verschillende statistische eigenschappen hebben. Zo zullen bijvoorbeeld clusters die bij een hoge sensitiviteit gedetecteerd zijn, groter zijn en meer gedetecteerde kandidaat-microcalcificaties bevatten. 'True positive' clusters die bij hoge sensitiviteit gedetecteerd zijn zouden zelfs een groot aantal 'false positive' gedetecteerde microcalcificaties kunnen bevatten. Dit verschijnsel limiteert het onderscheidend vermogen van kenmerken die individuele microcalcificaties beschrijven. Kenmerken die lijn- en randstructuren alsmede de afstand tussen microcalcificaties beschrijven bleken belangrijk te zijn met betrekking tot de onderzochte niveaus van sensitiviteit. Verder werd duidelijk dat de pixelwaarde van de achtergrond, als aanwijzing dat een cluster in een vetrijk gebied ligt of in klierrijk weefsel, een belangrijke waarde te hebben bij het classificeren. Bij hoge sensitiviteit gaf uiteindelijk de clusteroppervlakte het grootste onderscheidend vermogen.

Tenslotte werd in **hoofdstuk 5** een volledig automatische methode beschreven die ontwikkeld is om microcalcificatieclusters te kunnen karakteriseren in benigne en maligne clusters. In eerste instantie werden microcalcificaties automatisch gedetecteerd door gebruik te maken van de statistische methode die beschreven werd in hoofdstuk 3. Een totaal aantal van zestien kenmerken werd in deze studie gebruikt. Verder werd het effect op het karakteriseren van drie andere (meer eenvoudige) segmentatiemethoden onderzocht. Ook werd het effect van een heuristische methode onderzocht, bedoeld voor het linken van corresponderende clusters tussen twee opnames van dezelfde borst. De linkinginformatie werd gebruikt in kenmerkberekening en bij het karakteriseren.

Tien radiologen beoordeelden mammogrammen uit een subset op een lichtkast en gaven een mate van maligniteit aan voor iedere patiënt. Alle deelnemers hadden tenminste vier jaar klinische mammografie-ervaring en ondergingen juist een training om screeningsradioloog te worden. Resultaten op deze subset gaven aan dat de diagnostische prestatie van de automatische methode aanzienlijk hoger uitviel dan de beoordeling van de radiologen. Met betrekking tot de grotere dataset suggereren de resultaten dat het beschreven systeem voor 'computer aided diagnosis' het aantal fout positieve doorverwijzingen met 20% kan reduceren bij bijna 100% sensitiviteit. Verder kwam naar voren dat de relatieve afstand tot de pectorale spier en de borstrand, de standaarddeviaties van microcalcificatie-oppervlakte, -oriëntatie en -contrast belangrijke kenmerken zijn. Het gaf echter een beter resultaat wanneer meer kenmerken gebruikt werden in de classificatiemethode. De heuristische methode voor het linken van clusters, bleek zowel bij het berekenen van kenmerken als bij het classificeren van waarde te zijn.

In de toekomst zou deze studie uitgebreid kunnen worden door de prestaties van een aantal ervaren screeningsradiologen met lange ervaring op het gebied van de mammadia-

gnostiek te vergelijken met het systeem. Verder zou de reproduceerbaarheid van de resultaten onderzocht kunnen worden door gebruik te maken van andere datasets. Ook is het van belang datasets te gebruiken die het hele spectrum representeren van microcalcificatieclusters zoals die in de screeningspopulatie voorkomen (inbegrepen de fout negatieve gevallen in de screening).

List of publications

Journal publications

W J H Veldkamp, N Karssemeijer,
Accurate Segmentation and Contrast Measurement of Microcalcifications in Mammograms: a Phantom Study
Medical Physics 25 (7), 102-111, 1998.

W J H Veldkamp, N Karssemeijer,
Normalization of local contrast in mammograms,
to appear in IEEE Transactions on Medical Imaging, 2000

W J H Veldkamp, N Karssemeijer, J D M Otten, J H C L Hendriks,
Classification of microcalcification clusters into malignant and benign types,
Submitted to Medical Physics.

Publications in conference proceedings

W J H Veldkamp, N Karssemeijer,
Influence of Segmentation on Classification of Microcalcifications in Digital Mammography
18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Amsterdam, The Netherlands, 1996

W J H Veldkamp, N Karssemeijer,
Automatic Classification of Clustered Microcalcifications in Digital Mammography,
in Digital Mammography Proceedings of the 3rd International Workshop on Digital Mammography, Chicago, IL, 9-12 June, edited by K Doi, M L Giger, R M Nishikawa and R A Schmidt (Elsevier, Amsterdam. 1996)

N Karssemeijer, G M te Brake, W J H Veldkamp,
Methods in Computer Aided Digital Mammography,
Medical Imaging Technology, Japanese Society of Medical Imaging Technology, Vol. 14, No.6 , 1996.

W J H Veldkamp, N Karssemeijer.
Improved Correction For Signal Dependent Noise
Applied to Automatic Detection of Microcalcifications
in Digital Mammography Proceedings of the 4th International Workshop on Digital Mammography,
Nijmegen, 7-10 June, edited by N Karssemeijer and M A O Thijssen, J H C L Hendriks, L J T O van Erning (Kluwer, Dordrecht. 1998)

W J H Veldkamp, N Karssemeijer.
An Improved Method for Detection of Microcalcifications Clusters in Digital Mammograms
Medical Imaging 1999: Image processing, K M Hanson (editor), Vol. 3661, pp 512-522, SPIE, 1999.

W J H Veldkamp, N Karssemeijer.
Fully Automated Classification of Microcalcification Cases
Referred from a Nation Wide Screening Program,
To appear in Digital Mammography Proceedings of the 5th International Workshop on Digital Mammography, Toronto, Canada, to be held 11-14 June, 2000

Miscellaneous publications

N Karssemeijer, W J H Veldkamp, G M te Brake, J H C L Hendriks,
Beoordeling van Mammogrammen met behulp van Neurale Netwerken.
Nederlands Tijdschrift voor Geneeskunde, 143(45), pp 2232-2236, 6 november 1999.

Dankwoord

Bij het tot stand komen van dit proefschrif zou ik niet zo maar voorbij kunnen gaan aan een aantal mensen dat mij de afgelopen jaren bijgestaan heeft.

Allereerst wil ik co-promotor Nico Karssemeijer bedanken voor zijn kritische doch opbouwende aanwijzingen en opmerkingen. Ik heb veel kunnen steunen op zijn grote kennis van zaken.

Vervolgens gaat mijn dank uit naar Jan Hendriks die als radioloog gespecialiseerd is in mammografie en screening naar borstkanker. Als co-promoter heeft hij voor mij een belangrijke brug kunnen slaan naar de medische achtergronden van het onderzoek. Verder kon hij de manuscripten vanuit zijn eigen discipline beoordelen en eventuele belangrijke aanvullingen voorstellen.

Zeer zeker wil ik ook promotor Stan Gielen noemen voor zijn begeleiding en belangrijke aanwijzingen en adviezen bij het voltooien van het proefschrift.

Hans Otten van de afdeling Epidemiologie wil ik bedanken voor zijn belangrijke bijdrage in het tot stand brengen van de dataset die zo belangrijk is geweest voor dit onderzoek. Ook zijn aanwijzingen met betrekking tot de inleiding en het laatste hoofdstuk in dit proefschrift heb ik zeer op prijs gesteld.

Ik zou dank willen uitspreken naar de radiologen die in training waren bij het Landelijk Referentiecentrum voor bevolkingsonderzoek op borstkanker (LRCB) en tijdens die periode mee hebben willen werken aan een van onze experimenten. Dit was van grote waarde bij het tot stand komen van het laatste hoofdstuk uit dit proefschrift.

Guido te Brake heeft mij zo'n vier en een halfjaar als collega-promovendus en ook als goede vriend van gezelschap voorzien. Dr. te Brake was zo vriendelijk om de introductie van dit proefschrift een aantal maal "ongemeen kritisch" onder de loep te nemen. Verder heb ik regelmatig een beroep mogen doen op zijn expertise met betrekking tot proefschriftvervaardiging.

Niels Cornelisse bracht met name buiten de werksfeer belangrijke afleiding. Memorabel zijn onder meer de zaalvoetbalduels bij FC Radboud die Niels, ik en de overige FC Radboud-spelers (onderwie ook Guido) iedere dinsdagavond op het scherpst van de snede uitvochten.

Mark, Henkjan, Mark en Bas, zorgden als collega's altijd weer voor de nodige discussiepunten tijdens de lunch en het nodige gezelschap elders buiten de werksfeer.

Ik wil mijn ouders, Marieke, Esther, Hans, Eva en mijn vrienden elders in het land noemen. Vaak zocht ik hen in de weekends op wat mij in de gelegenheid stelde om letterlijk en figuurlijk zo nu en dan volledig afstand te nemen van het onderzoek. Tenslotte waren mijn ouders altijd in staat mij te motiveren als dat nodig was.

Curriculum Vitae

Wouter Veldkamp (geboren op 25 februari 1970 te Voorschoten) bezocht van 1982 tot 1988 het Rijnlands Lyceum in Oegstgeest. In laatst genoemd jaar legde hij aan deze school met goed gevolg het VWO-eindexamen af. Aansluitend begon hij met de studie Elektrotechniek aan de Technische Universiteit Delft.

Een stage bij het AMC op de afdeling Electrofysiologie bracht de auteur van dit proefschrift in contact met onderzoek in een medische omgeving. De maatschappelijke relevantie van dergelijk onderzoek sprak hem aan. Dit bleek ook een inspirerende factor bij de afstudeeropdracht die onder de vlag van de groep Verkeersbegeleidingssystemen aan de faculteit Elektrotechniek plaats vond.

Na in 1994 afgestudeerd te zijn had de auteur een korte doch leerzame militaire-dienstervaring. Vervolgens ging hij in april 1995 aan de slag als AIO in Nijmegen. Onder leiding van dr.ir. Nico Karssemeijer begon hij bij de afdeling Radiologie van het Radboud ziekenhuis te werken aan een automatische methode om clusters van microcalcificaties in mammogrammen te beoordelen. Hierbij kwam de opgedane ervaring met classificatiemethoden tijdens de eerder genoemde afstudeeropdracht goed van pas. Het promotieonderzoek heeft uiteindelijk in dit proefschrift mogen resulteren.

Sinds 15 april 1999 heeft de auteur een aanstelling voor drie jaar aan de afdeling Radiologie als onderzoeker en als aspirant klinisch fysicus.

R2 Technology, Inc, is proud to sponsor this work as part of its dedication to the improvement of mammography practice by assisting in the detection of early stage breast cancer.